GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 7, 2003, 18:46:24 ; Search time 19 Seconds
(without alignments)
597.258 Million cell updates/sec

Title: US-09-898-860-2
Perfect score: 649
Sequence: 1 MPREDAHFIYGYPKKGHGHS.....NAPPAYEKLSAEQSPPPYSP 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	649	100.0	118	2	A55253	melanoma antigen M
2	87	13.4	4957	2	T03455	ALR protein - huma
3	87	13.4	5262	2	T03454	ALR protein - huma
4	80.5	12.4	344	2	B28967	T-cell surface gly
5	80.5	12.4	344	2	I49585	CD2 antigen protei
6	79	12.2	972	2	T49773	related to actin-i
7	77	11.9	710	2	E69665	nitrate reductase
8	73	11.2	1069	2	T00043	BH-protocadherin-a
9	72.5	11.2	1429	2	T13720	gene expanded prot
10	71.5	11.0	1265	2	T51314	probable CO-induce
11	71	10.9	1072	2	T00041	BH-protocadherin P
12	71	10.9	1200	2	T00042	BH-protocadherin P
13	70.5	10.9	254	2	H83334	probable transcrip
14	70	10.8	335	2	G88640	protein F52C12.3 [
15	70	10.8	346	2	H84512	hypothetical prote
16	70	10.8	2100	2	T03223	probable polyketid
17	69.5	10.7	341	2	T22633	hypothetical prote
18	69.5	10.7	404	2	H64175	hypothetical prote
19	69	10.6	940	2	S19702	fibronectin-bindin
20	68.5	10.6	384	2	E72100	hypothetical prote
21	68.5	10.6	476	2	T35769	probable transmemb
22	68	10.5	316	2	C82085	conserved hypothet
23	68	10.5	377	2	C91270	beta-lactamase [im
24	68	10.5	377	2	C86111	beta-lactamase, pe
25	67.5	10.4	188	2	A64332	probable phosphono
26	67.5	10.4	315	2	T47971	seven in absentia-
27	67.5	10.4	476	2	T43464	hypothetical prote
28	67.5	10.4	670	2	T13739	probable hormone r
29	67.5	10.4	753	2	JQ0532	OP protein - Kenne
30	67	10.3	485	2	S32171	hydroxyneurosporen
31	67	10.3	2723	2	T03221	probable polyketid
32	66.5	10.2	215	2	F71923	hypothetical prote
33	66	10.2	140	2	S67666	probable membrane
34	66	10.2	751	2	T31515	hypothetical prote
35	65.5	10.1	258	2	A48820	homeobox protein (
36	65.5	10.1	384	2	A86521	hypothetical prote
37	65.5	10.1	518	2	D69539	conserved hypothet
38	65.5	10.1	1023	2	T48997	epsin-like protein
39	65.5	10.1	1436	2	A46496	antigen WC1.1 prec
40	65	10.0	309	1	E65112	hypothetical 34.6
41	65	10.0	309	2	B91140	hypothetical prote
42	65	10.0	309	2	E85985	hypothetical prote
43	65	10.0	309	2	AH0906	conserved hypothet
44	64.5	9.9	243	2	JE0204	testicular protein
45	64.5	9.9	359	2	A22891	fixZ protein - Rhi

ALIGNMENTS

RESULT 1
A55253
melanoma antigen MART-1 - human
N;Alternate names: melan-A protein
C;Species: Homo sapiens (man)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 04-Mar-2000
C;Accession: A55253; I38506
R;Kawakami, Y.; Eliyahu, S.; Delgado, C.H.; Robbins, P.F.; Rivoltini, L.; Topalian
Proc. Natl. Acad. Sci. U.S.A. 91, 3515-3519, 1994
A;Title: Cloning of the gene coding for a shared human melanoma antigen recognized
A;Reference number: A55253; MUID:94224770; PMID:8170938
A;Accession: A55253
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-118 <KAW>
A;Cross-references: GB:U06452; NID:g476131; PIDN:AAA19238.1; PID:g476132
R;Coulie, P.G.; Brichard, V.; Van Pel, A.; Wolfel, T.; Schneider, J.; Traversari, C
J. Exp. Med. 180, 35-42, 1994
A;Title: A new gene coding for a differentiation antigen recognized by autologous
A;Reference number: I38506; MUID:94275389; PMID:8006593
A;Accession: I38506
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-118 <RES>
A;Cross-references: EMBL:U06654; NID:g517022; PIDN:AAA20389.1; PID:g517023
C;Genetics:
A;Gene: GDB:MLANA
A;Cross-references: GDB:358979
A;Map position: 17q21-17q24
C;Superfamily: human melanoma antigen MART-1

Query Match 100.0%; Score 649; DB 2; Length 118;
Best Local Similarity 100.0%; Pred. No. 3.5e-61;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
```

RESULT 2
T03455
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T03455
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.;
Oncogene 15, 549-560, 1997

A;Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A;Reference number: Z14954; MUID:97388474; PMID:9247308
A;Accession: T03455
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4957 <PRA>
A;Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: human ALR protein
C;Keywords: alternative splicing

Query Match 13.4%; Score 87; DB 2; Length 4957;
Best Local Similarity 31.2%; Pred. No. 3.2;
Matches 25; Conservative 7; Mismatches 30; Indels 18; Gaps 3;

Qy 41 LLIGCWYCRRRNGYRALMDKSLHVGTQCALTR----RCPQEGFDHRDSKVSLQEKNCEPV 96
Db 939 LLIQCRHCER----------WMHAGCESLFTEDDVDHAPDEGFD----CVSCQPYVVKPV 984

Qy 97 VPNAPPAYEKLSAEQSPPPY 116
Db 985 APVAPPELVPMKVKEPEPQY 1004

RESULT 3
T03454
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T03454
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A;Reference number: Z14954; MUID:97388474; PMID:9247308
A;Accession: T03454
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5262 <PRA>
A;Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: human ALR protein
C;Keywords: alternative splicing

Query Match 13.4%; Score 87; DB 2; Length 5262;
Best Local Similarity 31.2%; Pred. No. 3.4;
Matches 25; Conservative 7; Mismatches 30; Indels 18; Gaps 3;

Qy 41 LLIGCWYCRRRNGYRALMDKSLHVGTQCALTR----RCPQEGFDHRDSKVSLQEKNCEPV 96
Db 1244 LLIQCRHCER----------WMHAGCESLFTEDDVDHAPDEGFD----CVSCQPYVVKPV 1289

Qy 97 VPNAPPAYEKLSAEQSPPPY 116
Db 1290 APVAPPELVPMKVKEPEPQY 1309

RESULT 4
B28967
T-cell surface glycoprotein CD2 precursor - mouse
N;Alternate names: CD2 antigen; T-lymphocyte antigen CD2; T11 protein
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1989 #sequence_revision 03-Jun-1993 #text_change 23-Jul-1999
C;Accession: B28967; S01347; S02293
R;Diamond, D.J.; Clayton, L.K.; Sayre, P.H.; Reinherz, E.L.
Proc. Natl. Acad. Sci. U.S.A. 85, 1615-1619, 1988
A;Title: Exon-intron organization and sequence comparison of human and murine T11 (CD2)
A;Reference number: A28967; MUID:88144486; PMID:2894031
A;Accession: B28967
A;Molecule type: mRNA
A;Residues: 1-344 <DIA>
A;Cross-references: GB:M19807; NID:g192479; PIDN:AAA37393.1; PID:g387122; GB:J03622
A;Note: the authors translated the codon TAT for residue 99 as Thr
R;Clayton, L.K.; Sayre, P.H.; Novotny, J.; Reinherz, E.L.
Eur. J. Immunol. 17, 1367-1370, 1987
A;Title: Murine and human T11 (CD2) cDNA sequences suggest a common signal transduc
A;Reference number: S01347; MUID:88004738; PMID:2820751
A;Accession: S01347
A;Molecule type: mRNA
A;Residues: 1-127,'M',129-174,'N',176-190,'NM',193-344 <CLA>
A;Cross-references: EMBL:X06143; NID:g54223; PIDN:CAA29500.1; PID:g54224
R;Sewell, W.A.; Brown, M.H.; Owen, M.J.; Fink, P.J.; Kozak, C.A.; Crumpton, M.J.
Eur. J. Immunol. 17, 1015-1020, 1987
A;Title: The murine homologue of the T lymphocyte CD2 antigen: molecular cloning, cl
A;Reference number: S02293; MUID:87276135; PMID:2440689
A;Accession: S02293
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-127,'M',129-174,'N',176-191,'M',193-344 <SEW>
A;Cross-references: EMBL:Y00023; NID:g50346; PIDN:CAA68258.1; PID:g50347
C;Genetics:
A;Map position: 3
C;Superfamily: T-cell surface glycoprotein CD2
C;Keywords: glycoprotein; surface antigen; T-cell; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-344/Product: T-cell surface glycoprotein CD2 #status predicted <MAT>
F;23-203/Domain: extracellular #status predicted <EXT>
F;204-228/Domain: transmembrane #status predicted <TMM>
F;229-344/Domain: intracellular #status predicted <INT>

Query Match 12.4%; Score 80.5; DB 2; Length 344;
Best Local Similarity 27.4%; Pred. No. 0.95;
Matches 31; Conservative 15; Mismatches 46; Indels 21; Gaps 5;

Qy 3 REDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSL 62
Db 188 KESKTEVVNCPEKGLSF-YVTVGVGAG-GLLLVLL-VALFIFC-ICKRRKRNRRRKDEEL 243

Qy 63 HVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPP 115
Db 244 EI----------KASRTSTVERGPKPHSTPAAAAQNSVALQAPPPP 279

RESULT 5
I49585
CD2 antigen protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 23-Jul-1999
C;Accession: I49585
R;Yagita, H.; Okumura, K.; Nakauchi, H.
J. Immunol. 140, 1321-1326, 1988
A;Title: Molecular cloning of the murine homologue of CD2: Homology of the molecule
A;Reference number: I49585; MUID:88140313; PMID:3257775
A;Accession: I49585
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-344 <RES>
A;Cross-references: GB:M18934; NID:g192486; PIDN:AAA37397.1; PID:g309158
C;Superfamily: T-cell surface glycoprotein CD2

Query Match 12.4%; Score 80.5; DB 2; Length 344;
Best Local Similarity 27.4%; Pred. No. 0.95;
Matches 31; Conservative 15; Mismatches 46; Indels 21; Gaps 5;

Qy 3 REDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSL 62
Db 188 KESKTEVVNCPEKGLSF-YVTVGVGAG-GLLLVLL-VALFIFC-ICKRRKRNRRRKDEEL 243

Qy 63 HVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPP 115
Db 244 EI----------KASRTSTVERGPKPHSTPAAAAQNSVALQAPPPP 279

```
RESULT 6
T49773
related to actin-interacting protein AIP3 [imported] - Neurospora crassa
N;Alternate names: protein B9J10.100
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49773
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49773
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-972 <SCH>
A;Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.100
A;Experimental source: BAC clone B9J10; strain OR74A
C;Genetics:
A;Gene: NCSP:B9J10.100
A;Map position: 6
A;Introns: 29/3; 161/1; 329/1

  Query Match             12.2%;  Score 79;  DB 2;  Length 972;
  Best Local Similarity   30.6%;  Pred. No. 4.1;
  Matches   26;  Conservative   10;  Mismatches   31;  Indels   18;  Gaps    4;

Qy         52 NGYR--ALMDKSLHVG-TQCALTRRCPQEGFDHRDSKVSLQE-------KNC-------- 93
              ||||  :  | | : | |  : ||| | :    | |  | |        :|
Db        165 NGYRPDSFRDDSRNSGPTTASPTRRLPSQRDQSRGSVTSDQSSLSSNTMQNIPVMAPYPG 224

Qy         94 EPVVPNAPPAYEKLSAEQSPPPYSP 118
              |  :|  |||  ::: :  |||  |
Db        225 EETIPTGPPAPPEINVDNFPPPPPP 249


RESULT 7
E69665
nitrate reductase (EC 1.7.99.4) nasC - Bacillus subtilis
N;Alternate names: assimilatory nitrate reductase; nitrate reductase (narB)
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: E69665; I40028
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seroi
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69665
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-710 <KUN>
A;Cross-references: GB:Z99105; GB:AL009126; NID:g2632457; PIDN:CAB12125.1; PID:g2632617
A;Experimental source: strain 168
R;Ogasawara, N.; Fujita, Y.; Kobayashi, Y.; Sadaie, Y.; Tanaka, T.; Takahashi, H.; Yaman
Microbiology 141, 257-259, 1995
A;Title: Systematic sequencing of the Bacillus subtilis genome: progress report of the J
A;Reference number: I39887; MUID:95219077; PMID:7704252
A;Accession: I40028
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 35-710 <RES>
A;Cross-references: GB:D30689; NID:g710016; PIDN:BAA06353.1; PID:g710019
C;Genetics:
A;Gene: nasC
C;Superfamily: formate dehydrogenase
C;Keywords: 4Fe-4S; chromoprotein; iron-sulfur protein; metalloprotein; molybdenum
F;26,29,33,63/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
F;154/Binding site: molybdopterin (Cys) (covalent) #status predicted

  Query Match             11.9%;  Score 77;  DB 2;  Length 710;
  Best Local Similarity   32.4%;  Pred. No. 4.8;
  Matches   34;  Conservative    8;  Mismatches   37;  Indels   26;  Gaps    6;

Qy         15 KGHGHSYTTAEE------AAGIGILTVILGVLLLIGCWY--CRRRNGYRALMDKSLHVGT 66
              ||   |||:||:       |  |      |:    |  |   ||  |      :| | ||
Db        507 KGRYFSYTSAEDIFNELREASRG------GIADYSGISYGRLRREGGIHWPCPESDHPGT 560

Qy         67 QCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPA-YEKLSAE 110
                  |     | | | | | :|       |:|| ||   || :|:
Db        561 GRLFT-----ESFAHPDQKAALS------VIPNEPPVPKEKPTAD 594


RESULT 8
T00043
BH-protocadherin-a - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C;Accession: T00043
R;Yoshida, K.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z14075
A;Accession: T00043
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1069 <YOS>
A;Cross-references: EMBL:AB006758; NID:d1227200; PIDN:BAA32597.1; PID:d1033562
C;Genetics:
A;Gene: pcdh7
A;Map position: 5C3-D

  Query Match             11.2%;  Score 73;  DB 2;  Length 1069;
  Best Local Similarity   25.6%;  Pred. No. 19;
  Matches   30;  Conservative   15;  Mismatches   32;  Indels   40;  Gaps    6;

Qy          9 IYGYPKKGHGHSYTTAEEAAGI------GILTVILGVLLLIGCWYCR--RRNGYRALMDK 60
              | |:|      ||  :::   |      ||:|||| :|:::   |||   :||| |
Db        865 IAGHP------SYEISKQRLSIVIGVVAGIMTVILIILIVMMARYCRSKNKNGYEA---- 914

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYS 117
                | :          || |      |          :  |  :| : :   | ||
Db        915 ----GKK-----------DHEDFFTPQQHDK------SKKPKKDKKNKKSKQPLYS 949


RESULT 9
T13720
gene expanded protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 17-Nov-2000
C;Accession: T13720
R;Boedigheimer, M.; Laughon, A.
Development 118, 1291-1301, 1993
A;Title: Expanded: a gene involved in the control of cell proliferation in imagina
A;Reference number: Z17740; MUID:94094747; PMID:8269855
A;Accession: T13720
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1429 <BOE>
A;Cross-references: EMBL:L14768; NID:g1763686; PID:g1763687; PIDN:AAB39774.1
C;Genetics:
A;Gene: expanded
A;Cross-references: FlyBase:FBgn0004583
C;Function:
A;Description: involved in control of cell proliferation in imaginal disk

  Query Match             11.2%;  Score 72.5;  DB 2;  Length 1429;
  Best Local Similarity   23.4%;  Pred. No. 30;
```

```
  Matches   37;  Conservative   10;  Mismatches   52;  Indels   59;  Gaps    7;

Qy          2 PREDAHFIYGYPKKGHG-------HSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGY 54
              || |::   |   :| |       ||  :|||              |::|      | |
Db        708 PRSDSNVSTGSSFRGDGSDPTDNKHSLLSAEELTN----------LIVG------RGTY 750

Qy         55 --RALMDKSLH------------VGTQCALTRRCPQEGFDHRDSKVSL------------ 88
                |  :  |||            | :       |   :  |  |  |
Db        751 PSRKTVSSSLHSDCDYVTLPLGDQGEEEVDQPPAPPPPYSARHEKTGLCGPPIAKPIPKP 810

Qy         89 --------QEKNCEPVVPNAP-PAYEKLSAEQSPPPYS 117
                      | | || || ||       : |||||
Db        811 IAVVAPKPDSPPCSPPVPPAPIPAPPPAIRRRDPPPYS 848


RESULT 10
T51314
probable CO-induced hydrogenase complex chain CooM [imported] - Rhodospirillum rubrum
C;Species: Rhodospirillum rubrum
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51314
R;Kerby, R.L.
submitted to the EMBL Data Library, July 1996
A;Reference number: Z25372
A;Accession: T51314
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1265 <KER>
A;Cross-references: EMBL:U65510; PIDN:AAC45116.1
A;Experimental source: strain UR1
C;Genetics:
A;Gene: cooM

  Query Match           11.0%;  Score 71.5;  DB 2;  Length 1265;
  Best Local Similarity  30.9%;  Pred. No. 33;
  Matches   25;  Conservative    8;  Mismatches   33;  Indels   15;  Gaps    3;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              :|  :| :: | |  |  |  |           ||: || : |||  ||  |:|||
Db       1172 LPVIEASWLGGLPGPGGWHPLT----------LTLALGAVGLIGWLYC--RDGYRHRAAS 1219

Qy         61 SLH---VGTQCALTRRCPQEG 78
              : |   |    |     | |
Db       1220 TTHSCGVSDIAASAMHVPASG 1240


RESULT 11
T00041
BH-protocadherin PCDH7 (clone BH-Pcdh-b) - human
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 01-Dec-2000
C;Accession: T00041; T00040
R;Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.; Sugano, S.
Genomics 49, 458-461, 1998
A;Title: Cloning, expression analysis, and chromosomal localization of BH-protocadherin
A;Reference number: Z14074; MUID:98277460; PMID:9615233
A;Accession: T00041
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1072 <YOS>
A;Cross-references: EMBL:AB006756; NID:g2979419; PIDN:BAA25195.1; PID:g2979420
A;Experimental source: clone BH-Pcdh-b
A;Accession: T00040
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1058,'M',1060,'LH',1063,'Y',1065,'TVFG' <YO2>
A;Cross-references: EMBL:AB006755; NID:g2979417; PIDN:BAA25194.1; PID:g2979418
A;Experimental source: clone BH-Pcdh-a
C;Genetics:
A;Map position: 4p15

  Query Match           10.9%;  Score 71;  DB 2;  Length 1072;
  Best Local Similarity  25.5%;  Pred. No. 31;
  Matches   27;  Conservative   14;  Mismatches   31;  Indels   34;  Gaps    5;

Qy         20 SYTTAEEAAGI------GILTVILGVLLLIGCWYCR--RRNGYRALMDKSLHVGTQCALT 71
              ||  :::   |      ||:|||| :|:::  |||   :||| |        | :
Db        870 SYEISKQRLSIVIGVVAGIMTVILIILIVVMARYCRSKNKNGYEA--------GKK---- 917

Qy         72 RRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYS 117
                     || |     |         :  |  :| : :   | ||
Db        918 --------DHEDFFTPQQHDK------SKKPKKDKKNKKSKQPLYS 949


RESULT 12
T00042
BH-protocadherin PCDH7 (clone BH-Pcdh-c) - human
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C;Accession: T00042
R;Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.; Sugano, S.
Genomics 49, 458-461, 1998
A;Title: Cloning, expression analysis, and chromosomal localization of BH-protocadhe
A;Reference number: Z14074; MUID:98277460; PMID:9615233
A;Accession: T00042
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1200 <YOS>
A;Cross-references: EMBL:AB006757; NID:g2979421; PIDN:BAA25196.1; PID:g2979422
A;Experimental source: clone BH-Pcdh-c
C;Genetics:
A;Map position: 4p15

  Query Match           10.9%;  Score 71;  DB 2;  Length 1200;
  Best Local Similarity  25.5%;  Pred. No. 35;
  Matches   27;  Conservative   14;  Mismatches   31;  Indels   34;  Gaps    5;

Qy         20 SYTTAEEAAGI------GILTVILGVLLLIGCWYCR--RRNGYRALMDKSLHVGTQCALT 71
              ||  :::   |      ||:|||| :|:::  |||   :||| |        | :
Db        823 SYEISKQRLSIVIGVVAGIMTVILIILIVVMARYCRSKNKNGYEA--------GKK---- 870

Qy         72 RRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYS 117
                     || |     |         :  |  :| : :   | ||
Db        871 --------DHEDFFTPQQHDK------SKKPKKDKKNKKSKQPLYS 902


RESULT 13
H83334
probable transcription regulator PA2488 [imported] - Pseudomonas aeruginosa (strain
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83334
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83334
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <STO>
A;Cross-references: GB:AE004676; GB:AE004091; NID:g9948532; PIDN:AAG05876.1; GSPDB:G
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2488

  Query Match           10.9%;  Score 70.5;  DB 2;  Length 254;
  Best Local Similarity  25.7%;  Pred. No. 7.9;
  Matches   27;  Conservative   11;  Mismatches   40;  Indels   27;  Gaps    5;

Qy          2 PREDAHFIYGYPKKGHGH-------SYTTAEEAAGI----GILTVILGVLLLI------- 43
              ||: |    |:|   | |         :|  ||    :    |:|     | |::
Db         53 PRQIAWIPPGHPHSAHSHGSLAGWSAYLAAEHCRDLPRHPGVLACSAFVALIVERATRWP 112

Qy         44 --GCWYCRRRNGYRALMDKSLHVGTQCALTRRCPQEGFDHRDSKV 86
```

```
                |   | |  | |:|:  |   |     | |  |    ||::
Db        113 TGSAWSTERENLLRVLLDELRHAVPQ---RTRLPYPG----DSRL 150
```

RESULT 14
G88640
protein F52C12.3 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: G88640
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: G88640
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-335 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAC68961.1; PID:g3800953; GSPDB:GN00022; CESP:F52C12
C;Genetics:
A;Gene: F52C12.3
A;Map position: 4

```
  Query Match             10.8%;  Score 70;  DB 2;  Length 335;
  Best Local Similarity   33.8%;  Pred. No. 12;
  Matches   23;  Conservative    5;  Mismatches   28;  Indels   12;  Gaps    3;

Qy         51 RNGYRALMDKSLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAE 110
              |||  ::  | ||  |  ||    | |  :   :   ||  | :|  || ||
Db        161 RNGQISMKMKVLH--TSPALVNFLPLEMLESIQTSTILQ--NLKPPTPNRPPC------- 209

Qy        111 QSPPPYSP 118
                || | |
Db        210 -RPPSYWP 216
```

RESULT 15
H84512
hypothetical protein At2g14000 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: H84512
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: H84512
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <STO>
A;Cross-references: GB:AE002093; NID:g4388823; PIDN:AAD19778.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g14000
A;Map position: 2

```
  Query Match             10.8%;  Score 70;  DB 2;  Length 346;
  Best Local Similarity   22.7%;  Pred. No. 12;
  Matches   30;  Conservative   19;  Mismatches   57;  Indels   26;  Gaps    5;

Qy          5 DAHFI--------YGYPKK-GHGHSYTTAEE----------AAGIGILTVILGVLLLIGC 45
              | | |        | ||    |   |    |           ||  |   :  :: |
Db        218 DEHVISNDEIYEDYAYPTSLAHAQPYLLLPEDPTPYRVSPPPAGDDISQQMAWMIAL--- 274

Qy         46 WYCRRRNGYRALMDKSLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYE 105
                |: |     | :::     |  |||   :  :  :   :::::: | :|| | |:
Db        275 --TRKNNSMMHKMWRAISRIRPCVCTRRGGVDDANRENQSKTIEDRHAERLVPAAGPS-- 330

Qy        106 KLSAEQSPPPYS 117
                |  | || :|
Db        331 GASTSQPPPEHS 342
```

Search completed: October 7, 2003, 18:49:41
Job time : 21 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 7, 2003, 18:39:49 ; Search time 41 Seconds
(without alignments)
456.823 Million cell updates/sec

Title: US-09-898-860-2
Perfect score: 649
Sequence: 1 MPREDAHFIYGYPKKGHGHS.....NAPPAYEKLSAEQSPPPYSP 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	649	100.0	118	15	AAR63158	Tumour rejection a
2	649	100.0	118	16	AAR84212	MART-1 melanoma an
3	649	100.0	118	20	AAY42634	Human melanoma ant
4	649	100.0	118	20	AAW83134	Human tumour rejec
5	649	100.0	118	22	AAU28888	MART-1 antigen. H
6	649	100.0	118	22	AAB86042	Human MART1/MelanA
7	649	100.0	118	23	AAU77793	Human melanoma ant
8	649	100.0	118	23	AAU84804	Human MART consens
9	649	100.0	118	23	AAU11541	Human Melanoma tum
10	649	100.0	119	20	AAY42633	Human melanoma ant
11	649	100.0	140	24	ABG76339	Recombinant mouse
12	649	100.0	496	24	ABG76343	Mouse recombinant
13	645	99.4	118	23	AAU77810	Human MART-1 prote
14	643	99.1	118	23	AAU77809	Human MART-1 prote
15	642	98.9	118	20	AAY31980	Human MART1 melano
16	641	98.8	118	23	AAU77806	Human MART-1 prote
17	641	98.8	118	23	AAU77808	Human MART-1 prote
18	641	98.8	118	23	AAU77811	Human MART-1 prote
19	640	98.6	118	23	AAU77807	Human MART-1 prote
20	599	92.3	118	23	AAU77815	Human MART-1 prote
21	598	92.1	118	23	AAU77813	Human MART-1 prote
22	598	92.1	118	23	AAU77814	Human MART-1 prote
23	597	92.0	118	23	AAU77812	Human MART-1 prote
24	596	91.8	118	23	AAU77816	Human MART-1 prote
25	562	86.6	104	23	AAU98926	Human melanoma ant
26	532	82.0	98	23	ABG70354	Novel human thromb
27	427.5	65.9	113	20	AAY31979	Mouse MART1 melano
28	427.5	65.9	114	20	AAY42632	Murine melanoma an
29	323	49.8	94	23	ABG70355	Novel human thromb
30	262.5	40.4	5546	23	AAU85008	Human melanocyte d
31	209	32.2	40	22	AAB61643	MART 1 peptide #1.
32	173	26.7	30	23	AAU84868	Human MART segment
33	169	26.0	30	23	AAU84870	Human MART segment
34	167	25.7	30	23	AAU84871	Human MART segment
35	163	25.1	30	23	AAU84869	Human MART segment
36	161	24.8	30	23	AAU84867	Human MART segment
37	158	24.3	30	23	AAU84865	Human MART segment
38	156	24.0	30	22	AAB61644	MART 1 peptide #2.
39	148	22.8	30	23	AAU84866	Human MART segment
40	121	18.6	23	23	ABG79128	Human MART-1 class
41	108	16.6	21	18	AAW00903	Human melanoma MAR
42	108	16.6	21	23	AAE20402	Human melanoma tum
43	98	15.1	22	23	AAU11546	Human control pept
44	82	12.6	17	23	AAU84872	Human MART segment
45	79	12.2	1114	22	ABB66628	Drosophila melanog

ALIGNMENTS

RESULT 1
AAR63158
ID AAR63158 standard; Protein; 118 AA.
XX
AC AAR63158;
XX
DT 25-MAR-2003 (updated)
DT 26-MAY-1995 (first entry)
XX
DE Tumour rejection antigen precursor.
XX
KW Tumour rejection antigen; precursor; HLA-A2 molecule; tyrosinase;
KW isolation; melanoma; cell line; LB-39-MEL; diagnosis; vaccine;
KW therapy.
XX
OS Homo sapiens.
XX
PN WO9421126-A1.
XX
PD 29-SEP-1994.
XX
PF 09-MAR-1994; 94WO-US02487.
XX
PR 18-MAR-1993; 93US-0032978.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Brichard V, De Plaen E, Traversari C;
PI Van Pel A, Wolfel T;
XX
DR WPI; 1994-316544/39.
DR N-PSDB; AAQ76370.

```
XX
PT   Nucleic acid coding for a tumour rejection antigen precursor - is
PT   used for developing prods. for diagnosis or treatment of expression
PT   related disorders, partic. melanoma
XX
PS   Claim 5; Page 14; 26pp; English.
XX
CC   This sequence represents the tumour rejection antigen precursor which is
CC   processed to a tumour rejection antigen presented by HLA-A2 molecules.
CC   The tumour rejection antigen is not related to tyrosinase.  The cDNA
CC   encoding this sequence was isolated from the melanoma cell line,
CC   LB-39-MEL.  The tumour rejection antigen may be used for diagnosis or
CC   in vaccines or for therapy of disorders characterised by the expression
CC   of the tumour rejection antigen precursor, particularly melanoma.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   118 AA;

  Query Match             100.0%;  Score 649;  DB 15;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-65;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 2
AAR84212
ID   AAR84212 standard; Protein; 118 AA.
XX
AC   AAR84212;
XX
DT   20-APR-1996  (first entry)
XX
DE   MART-1 melanoma antigen.
XX
KW   MART-1; melanoma antigen recognised by T-cell; melanoma;
KW   metastatic melanoma; tumour-associated antigen; immunogen;
KW   diagnosis; prognosis; prophylaxis; therapy; vaccine.
XX
OS   Mammalian.
XX
FH   Key             Location/Qualifiers
FT   Region          27..47
FT                   /note= "hydrophobic region"
XX
PN   WO9529193-A2.
XX
PD   02-NOV-1995.
XX
PF   21-APR-1995;   95WO-US05063.
XX
PR   05-APR-1995;   95US-0417174.
PR   22-APR-1994;   94US-0231565.
XX
PA   (USSH ) US SEC DEPT HEALTH.
XX
PI   Kawakami Y,  Rosenberg SA;
XX
DR   WPI; 1995-382963/49.
DR   N-PSDB; AAT02714.
XX
PT   DNA encoding melanoma antigens recognised by T-lymphocytes - also
PT   vectors, host cells and antibodies, used to detect, treat and
PT   immunise animal against melanoma.
XX
PS   Claim 11; Page 117; 184pp; English.
XX
CC   The melanoma antigen (MART-1) is produced by recombinant DNA
CC   methods, i.e. preferably using a baculovirus vector for expression
CC   in insect cell cultures.  MART-1 protein is a source of immunogenic
CC   peptides (see AAR84196 for peptide M9-2) which are optionally modified
CC   (see AAR84783-R84800) and used in medicaments for the treatment or
CC   prevention (by immunization) of melanoma.  Antibodies against MART-1
CC   and its immunogenic peptides may be used in the detection and
CC   isolation of MART-1 from a sample, the detection of which is
CC   indicative of a disease state (melanoma or metastatic melanoma).
XX
SQ   Sequence   118 AA;

  Query Match             100.0%;  Score 649;  DB 16;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-65;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 3
AAY42634
ID   AAY42634 standard; Protein; 118 AA.
XX
AC   AAY42634;
XX
DT   10-JAN-2000  (first entry)
XX
DE   Human melanoma antigen hMART1.
XX
KW   Immune response; self-antigen; immune effector cell; cancer; melanoma;
KW   mouse; melanoma antigen; MART1.
XX
OS   Homo sapiens.
XX
PN   WO9946988-A1.
XX
PD   23-SEP-1999.
XX
PF   19-MAR-1999;   99WO-US06034.
XX
PR   20-MAR-1998;   98US-0078890.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Nicolette CA;
XX
DR   WPI; 1999-580277/49.
DR   N-PSDB; AAZ07987.
XX
PT   Method of inducing an immune reaction to a self-antigen by
PT   administering the antigen, especially useful for treating cancer or
PT   melanoma  -
XX
PS   Disclosure; Fig 3A-B; 70pp; English.
XX
CC   The invention provides a method of inducing a prophylactic immune
CC   response to a self-antigen in a subject. The method comprises
CC   administering the antigen or its derivative or administering educated
CC   immune effector cells able to recognize and lyze cells expressing the
CC   self-antigen or its derivative. The method is used to stimulate an immune
CC   response against a self-antigen especially one expressed in a cancer or
CC   melanoma. The present sequence represents the human melanoma antigen
CC   hMART1.
XX
SQ   Sequence   118 AA;

  Query Match             100.0%;  Score 649;  DB 20;  Length 118;
```

```
  Best Local Similarity   100.0%;  Pred. No. 2.1e-65;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 4
AAW83134
ID   AAW83134 standard; Protein; 118 AA.
XX
AC   AAW83134;
XX
DT   04-FEB-1999  (first entry)
XX
DE   Human tumour rejection antigen precursor.
XX
KW   Human; tumour rejection antigen precursor; human leukocyte antigen;
KW   TRAP; HLA; cancer; melanoma.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "encoded by CGA"
FT   Misc-difference 17
FT                   /note= "encoded by GAC"
XX
PN   US5837476-A.
XX
PD   17-NOV-1998.
XX
PF   16-JAN-1998;   98US-0007966.
XX
PR   03-MAR-1995;   95US-0398409.
PR   16-JAN-1998;   98US-0007966.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Brichard V,  De Plaen E,  Traversari C;
PI   Van Pel A,  Woelfelt;
XX
DR   WPI; 1999-043967/04.
DR   N-PSDB; AAV70150.
XX
PT   Use of a tumour rejection antigen precursor - as a marker for
PT   diagnosing a disorder characterised by expression of a tumour
PT   rejection antigen precursor which is not tyrosinase
XX
PS   Claim 1; Column 7-9; 11pp; English.
XX
CC   A method has been developed for the diagnosis of a disorder which is
CC   characterised by the expression of a tumour rejection antigen precursor
CC   (TRAP) which is not tyrosinase, and which is processed to a TRA which
CC   forms a complex with an HLA-A2 molecule. The present sequence represents
CC   the TRAP for use in the present invention. The method comprises
CC   contacting a sample from a subject with an agent specific for the
CC   complex and determining the interaction between the complex and the
CC   agent as a determination of the disorder. TRAP can be used for the
CC   diagnosis and treatment of disorders characterised by the expression
CC   of the TRAP molecules such as cancers, particularly melanoma.
XX
SQ   Sequence   118 AA;

  Query Match             100.0%;  Score 649;  DB 20;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-65;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 5
AAU28888
ID   AAU28888 standard; Protein; 118 AA.
XX
AC   AAU28888;
XX
DT   18-DEC-2001  (first entry)
XX
DE   MART-1 antigen.
XX
KW   Human; MART-I; immunogenic; melanoma antigen recognised by T lymphocyte;
KW   diagnostic; therapeutic; vaccine; melanoma; in vivo tumour recognition;
KW   in vivo tumour rejection.
XX
OS   Homo sapiens.
XX
PN   US6270778-B1.
XX
PD   07-AUG-2001.
XX
PF   12-MAR-1999;   99US-0267439.
XX
PR   05-MAY-1998;   98US-0073138.
PR   22-APR-1994;   94US-0231565.
PR   05-APR-1995;   95US-0417174.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kawakami Y,  Rosenberg SA;
XX
DR   WPI; 2001-595403/67.
DR   N-PSDB; AAS45524.
XX
PT   Immunogenic peptide useful in vaccines comprises specific amino acids
PT   of new melanoma antigen recognised by T lymphocytes -
XX
PS   Claim 2 ; Figure 1; 73pp; English.
XX
CC   The invention relates to a novel immunogenic peptide comprising 5-20
CC   contiguous amino acids of new melanoma antigen recognised by T
CC   lymphocytes (MART-1). The peptide sequence contains at least one amino
CC   acid modification of MART-1. The peptide is used in diagnostic and
CC   therapeutic methods as an immunogen or vaccine to prevent or treat
CC   melanoma, and for in vivo tumour recognition and rejection. AAU28888-
CC   AAU29008 represent MART-1 peptide amino acid sequences, and related
CC   sequences of the invention.
XX
SQ   Sequence   118 AA;

  Query Match             100.0%;  Score 649;  DB 22;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-65;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 6
AAB86042
```

ID AAB86042 standard; Protein; 118 AA.
XX
AC AAB86042;
XX
DT 16-JUL-2001 (first entry)
XX
DE Human MART1/MelanA protein.
XX
KW Listeria; expression vector; tumor-associated antigen; Trp 1; Trp 2;
KW MelanA/MART-1; cytostatic; attenuated; immunotherapy; malignant melanoma;
KW pigmented tumor; malignant schwannoma; vaccination; tyrosinase;
KW antigen-presenting cell.
XX
OS Homo sapiens.
XX
PN WO200127295-A1.
XX
PD 19-APR-2001.
XX
PF 13-OCT-2000; 2000WO-DE03629.
XX
PR 14-OCT-1999; 99DE-1049594.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Schadendorf D, Paschen A, Chakraborty T, Domann E;
XX
DR WPI; 2001-282041/29.
DR N-PSDB; AAF88044.
XX
PT Listeria expression vector for immunotherapy, particularly of malignant
PT melanoma, comprises a DNA sequence encoding tumor-associated antigens -
XX
PS Disclosure; Fig 4; 41pp; German.
XX
CC This invention describes a novel Listeria expression vector (A) for
CC immunotherapy which comprises a promoter (P), functional in Listeria,
CC operably linked to a DNA sequence (I) encoding one of the
CC tumor-associated antigens (II) human tyrosinase, Trp 1 or 2, or
CC MelanA/MART-1. The products of the invention have cytostatic activity.
CC Recombinant attenuated Listeria containing (A) are useful for
CC immunotherapy (prophylactic, adjuvant or therapeutic), specifically of
CC malignant melanoma (but also other pigmented tumors such as malignant
CC schwannoma), particularly as a replacement for radiotherapy. Using
CC attenuated Listeria as carrier for (A) provides a simple way of
CC vaccination, since antigen-presenting cells acquire tumor-associated
CC antigens by natural infection, eliminating the need for labor-intensive
CC ex vivo modification of autologous cells. This sequence represents the
CC human MART-1/MelanA protein described in the method of the invention.
XX
SQ Sequence 118 AA;

Query Match 100.0%; Score 649; DB 22; Length 118;
Best Local Similarity 100.0%; Pred. No. 2.1e-65;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy 61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118

RESULT 7
AAU77793
ID AAU77793 standard; Protein; 118 AA.
XX
AC AAU77793;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human melanoma antigen recognised by T cells (MART-1).
XX
KW MART-1; melanoma antigen recognised by T cells; human;
KW anti-melanoma peptide; major histocompatibility complex; MHC;
KW immunotherapy; cancer; vaccine; immunoregulatory.
XX
OS Homo sapiens.
XX
PN WO200212272-A2.
XX
PD 14-FEB-2002.
XX
PF 03-AUG-2001; 2001WO-US24328.
XX
PR 04-AUG-2000; 2000US-223641P.
PR 13-DEC-2000; 2000US-255502P.
PR 25-JAN-2001; 2001US-264432P.
PR 26-MAR-2001; 2001US-279005P.
XX
PA (GENZ) GENZYME CORP.
XX
PI Nicolette CA;
XX
DR WPI; 2002-257459/30.
DR N-PSDB; ABK11763.
XX
PT Novel anti-melanoma peptide compounds useful for inducing immune
PT response in a subject, and in the preparation of medicaments for the
PT treatment and diagnosis of cancer -
XX
PS Disclosure; Page 69-70; 79pp; English.
XX
CC This invention relates to a novel anti-melanoma peptide compound
CC comprising a peptide of the human melanoma antigen recognised by T cells
CC (MART-1) protein. These compounds are designed to enhance binding to
CC major histocompatibility (MHC) compounds and enhance immunoregulatory
CC properties and induce an immune response. The invention also comprises
CC the nucleotide sequences encoding the peptides of the invention. The
CC compounds of the invention are useful for inducing an immune response in
CC a subject, by delivering the compounds in the context of a major
CC histocompatibility (MHC) molecule which presents the compound on the
CC surface of an antigen presenting cell, or by delivering it as a
CC polynucleotide that encodes the peptide. The invention also comprises
CC antibodies that recognise and bind these compounds which are useful in
CC immunotherapy. The compounds of the invention are useful for modulating
CC an immune response to synthetic and naturally occurring compounds in a
CC subject. The compounds may also be used as components of anti-cancer
CC vaccines and to expand immune effector cells that are specific for
CC cancers characterised by expression of the human melanoma antigen
CC recognised by T cells, MART-1. The compounds of the invention are also
CC useful for the detection and purification of antibodies and may be used
CC for the preparation of medicaments for the diagnosis and treatment of
CC diseases such as cancer. The compounds of the invention have enhanced
CC binding to MHC molecules and enhanced immunoregulatory properties
CC relative to their natural counterparts. The present sequence represents
CC the human melanoma antigen recognised by T cells (MART-1) used to create
CC the peptide compounds of the invention.
XX
SQ Sequence 118 AA;

Query Match 100.0%; Score 649; DB 23; Length 118;
Best Local Similarity 100.0%; Pred. No. 2.1e-65;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy 61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118

```
RESULT 8
AAU84804
ID   AAU84804 standard; Protein; 118 AA.
XX
AC   AAU84804;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human MART consensus sequence.
XX
KW   Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
KW   viral infection; human immunodeficieny virus; melanoma;
KW   bacterial infection; Salmonella; Legionella; parasitic infection;
KW   Trypanosoma; Toxoplasma; Giardia.
XX
OS   Homo sapiens.
XX
PN   WO200190197-A1.
XX
PD   29-NOV-2001.
XX
PF   25-MAY-2001; 2001WO-AU00622.
XX
PR   26-MAY-2000; 2000AU-0007761.
XX
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Thomson SA,  Ramshaw IA;
XX
DR   WPI; 2002-147575/19.
XX
PT   New synthetic polypeptides having several different segments of at
PT   least one parent polypeptide linked together differently compared to
PT   the linkage in the parent polypeptide, for inducing immune response
PT   against a pathogen or cancer -
XX
PS   Example 3; Fig 27; 364pp; English.
XX
CC   The invention relates to a new synthetic polypeptide (I) comprising
CC   several different segments of at least one parent polypeptide linked
CC   together in a different relationship relative to their linkage in the
CC   parent polypeptide to impede, abrogate or otherwise alter at least one
CC   function associated with the parent polypeptide and for inducing an
CC   immune response against a pathogen or cancer. Also included are a
CC   synthetic polynucleotide encoding and a computer system for
CC   designing the synthetic polypeptides. The synthetic polypeptides and
CC   polynucleotides are referred to as a Savine. The synthetic polypeptide is
CC   useful for modulating immune responses preferably directed against a
CC   pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
CC   colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
CC   liver, oesophagus, brain, testicle, uterus), as potentiating agents.
CC   Compositions comprising the polypeptide may be used in the treatment or
CC   prophylaxis against viral (such as infections caused by HIV (human
CC   immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
CC   virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
CC   (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
CC   Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
CC   (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
CC   Trypanosoma, Toxoplasma and Giardia) infections. The present
CC   sequence is a consensus sequence for a parent protein used to design a
CC   savine of the invention.
XX
SQ   Sequence   118 AA;

  Query Match           100.0%;  Score 649;  DB 23;  Length 118;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-65;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 9
AAU11541
ID   AAU11541 standard; protein; 118 AA.
XX
AC   AAU11541;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Human Melanoma tumour associated antigen.
XX
KW   Human; melanoma tumour associated antigen; human leukocyte antigen;
KW   major histocompatibility complex; MHC; HLA-A2.2; vaccine; cancer;
KW   HIV; human immunodeficiency virus infection; cytostatic; virucide;
KW   housekeeping epitope; adoptive immunotherapy; neoplastic disease;
KW   viral disease; hepatitis virus; papilloma virus; tumour; leukaemia;
KW   lymphoma; breast cancer; prostate cancer; lung cancer; melan-A;
KW   parasitic infection; Chlamydia; Trypanosoma; Toxoplasma.
XX
OS   Homo sapiens.
XX
PN   WO200182963-A2.
XX
PD   08-NOV-2001.
XX
PF   27-APR-2001; 2001WO-US13806.
XX
PR   28-APR-2000; 2000US-0560465.
PR   28-APR-2000; 2000US-0561074.
PR   28-APR-2000; 2000US-0561571.
PR   28-APR-2000; 2000US-0561572.
XX
PA   (CTLI-) CTL IMMUNOTHERAPIES CORP.
XX
PI   Simard JJL,  Diamond DC,  Lei X;
XX
DR   WPI; 2002-066492/09.
XX
PT   Novel vaccine useful for treating neoplastic and viral diseases,
PT   comprises a first housekeeping epitope derived from a first antigen
PT   associated with a first target cell -
XX
PS   Example 21; Fig 15; 131pp; English.
XX
CC   The invention relates to a vaccine comprising a first housekeeping
CC   epitope derived from a first antigen associated with a first target
CC   cell. Also included are an isolated T cell expressing a T cell receptor
CC   specific for a major histocompatibility complex (MHC)-peptide complex
CC   comprising a first housekeeping epitope which is derived from a first
CC   antigen associated with a first target cell, selecting an epitope
CC   (or peptide sequence) from a population of peptide fragments of an
CC   antigen associated with a target in a host, where the fragments have a
CC   known or predicted affinity for a MHC receptor peptide binding cleft of
CC   the host, where the epitope selected corresponds to a product of
CC   proteolytic cleavage of the antigen in a cell of the host and
CC   a nucleic acid construct comprising a first coding region, where the
CC   first coding region comprises a first sequence encoding at least a first
CC   polypeptide, where the first polypeptide comprises a first housekeeping
CC   epitope derived from a first antigen associated with a first target cell;
CC   The epitopes, peptides, vaccines and nucleic acids are useful in the
CC   manufacture of a medicament for use in adoptive immunotherapy and for
CC   prevention and treatment of neoplastic and viral diseases (e.g.
CC   human immunodeficiency virus, HIV, infection, hepatitis virus and
CC   papilloma virus), cancers (e.g. tumours, leukaemia, lymphoma, breast
CC   cancer, prostate cancer and lung cancer), infection of cells by
CC   intracellular parasites (e.g. Chlamydia, Trypanosoma and
CC   Toxoplasma) and many other examples given in the specification.
CC   The invention permits the vaccine designer to ignore peptides that,
CC   despite predicted high binding affinity for MHC, will never be useful
CC   because they cannot be presented by target cells. The invention provides
```

CC a major advance in vaccine design, one that combines the power of antigen
CC sequence analysis with the fundamental realities of immunology. The
CC invention allows for the simple and effective selection of meaningful
CC epitopes for creation of MHC class I or Class II vaccines using any
CC polypeptide sequence corresponding to a desired target. The present
CC sequence is an HLA-A2.1 (human leukocyte antigen) presenting target cell
CC protein from which epitopes of the invention may be derived, Mela-A
CC (melanoma tumour associated antigen).
XX
SQ Sequence 118 AA;

Query Match 100.0%; Score 649; DB 23; Length 118;
Best Local Similarity 100.0%; Pred. No. 2.1e-65;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
Db 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy 61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
Db 61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118

RESULT 10
AAY42633
ID AAY42633 standard; Protein; 119 AA.
XX
AC AAY42633;
XX
DT 10-JAN-2000 (first entry)
XX
DE Human melanoma antigen hMART1.
XX
KW Immune response; self-antigen; immune effector cell; cancer; melanoma;
KW mouse; melanoma antigen; MART1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 119
FT /note= "unspecified"
XX
PN WO9946988-A1.
XX
PD 23-SEP-1999.
XX
PF 19-MAR-1999; 99WO-US06034.
XX
PR 20-MAR-1998; 98US-0078890.
XX
PA (GENZ) GENZYME CORP.
XX
PI Nicolette CA;
XX
DR WPI; 1999-580277/49.
DR N-PSDB; AAZ07986.
XX
PT Method of inducing an immune reaction to a self-antigen by
PT administering the antigen, especially useful for treating cancer or
PT melanoma -
XX
PS Disclosure; Fig 2B; 70pp; English.
XX
CC The invention provides a method of inducing a prophylactic immune
CC response to a self-antigen in a subject. The method comprises
CC administering the antigen or its derivative or administering educated
CC immune effector cells able to recognize and lyze cells expressing the
CC self-antigen or its derivative. The method is used to stimulate an immune
CC response against a self-antigen especially one expressed in a cancer or
CC melanoma. The present sequence represents the human melanoma antigen
CC hMART1.
XX
SQ Sequence 119 AA;

Query Match 100.0%; Score 649; DB 20; Length 119;
Best Local Similarity 100.0%; Pred. No. 2.2e-65;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
Db 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy 61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
Db 61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118

RESULT 11
ABG76339
ID ABG76339 standard; Protein; 140 AA.
XX
AC ABG76339;
XX
DT 10-MAY-2003 (first entry)
XX
DE Recombinant mouse protein, MelanA/MART1.
XX
KW Mouse; protein targeting; exosome; lactadherin; C1 domain; C2 domain;
KW membrane vesicle; mutant; mutein.
XX
OS Mus sp.
OS Synthetic.
XX
PN WO2003016522-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-EP09108.
XX
PR 17-AUG-2001; 2001US-313159P.
PR 26-DEC-2001; 2001US-343991P.
XX
PA (ANOS-) ANOSYS INC.
XX
PI Delcayre A, Le Pecq J;
XX
DR WPI; 2003-268331/26.
XX
PT Targeting polypeptides to exosomes providing a chimeric genetic
PT construct and introducing the construct into exosome-producing cells in
PT vivo or ex vivo -
XX
PS Example 6; Page 85-86; 94pp; English.
XX
CC The present invention relates to a method and compounds for targeting
CC polypeptides to exosomes. The method comprises providing a chimeric
CC genetic construct encoding the polypeptide fused to a targeting
CC polypeptide comprising lactadherin or its portion comprising
CC a functional C1 and/or C2 domain, and introducing the construct into
CC exosome-producing cells in vivo or ex vivo, to generate recombinant
CC exosomes. The method is useful for targeting proteins to membrane
CC vesicles, particularly exosomes, and is useful in experimental,
CC research, therapeutic, prophylactic, and diagnostic areas. The
CC present sequence represents a recombinant mouse protein.
XX
SQ Sequence 140 AA;

Query Match 100.0%; Score 649; DB 24; Length 140;
Best Local Similarity 100.0%; Pred. No. 2.6e-65;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
Db 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

```
Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 12
ABG76343
ID   ABG76343 standard; Protein; 496 AA.
XX
AC   ABG76343;
XX
DT   10-MAY-2003  (first entry)
XX
DE   Mouse recombinant chimeric fusion protein, MART1/CCR7.
XX
KW   Mouse; protein targeting; exosome; lactadherin; C1 domain; C2 domain;
KW   membrane vesicle; mutant; mutein.
XX
OS   Mus sp.
OS   Synthetic.
XX
PN   WO2003016522-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-EP09108.
XX
PR   17-AUG-2001; 2001US-313159P.
PR   26-DEC-2001; 2001US-343991P.
XX
PA   (ANOS-) ANOSYS INC.
XX
PI   Delcayre A,  Le Pecq J;
XX
DR   WPI; 2003-268331/26.
XX
PT   Targeting polypeptides to exosomes providing a chimeric genetic
PT   construct and introducing the construct into exosome-producing cells in
PT   vivo or ex vivo  -
XX
PS   Claim 28; Page 90-92; 94pp; English.
XX
CC   The present invention relates to a method and compounds for targeting
CC   polypeptides to exosomes. The method comprises providing a chimeric
CC   genetic construct encoding the polypeptide fused to a targeting
CC   polypeptide comprising lactadherin or its portion comprising
CC   a functional C1 and/or C2 domain, and introducing the construct into
CC   exosome-producing cells in vivo or ex vivo, to generate recombinant
CC   exosomes. The method is useful for targeting proteins to membrane
CC   vesicles, particularly exosomes, and is useful in experimental,
CC   research, therapeutic, prophylactic, and diagnostic areas. The
CC   present sequence represents a mouse recombinant chimeric fusion
CC   protein.
XX
SQ   Sequence  496 AA;

  Query Match             100.0%;  Score 649;  DB 24;  Length 496;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-64;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 13
AAU77810
ID   AAU77810 standard; Protein; 118 AA.
XX
AC   AAU77810;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Human MART-1 protein mutant #5.
XX
KW   MART-1; melanoma antigen recognised by T cells; human; mutant;
KW   anti-melanoma peptide; major histocompatibility complex; MHC;
KW   immunotherapy; cancer; vaccine; immunoregulatory; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 27
FT                   /note= "Wild type A substituted by T"
FT   Misc-difference 28
FT                   /note= "Wild type A substituted by A"
XX
PN   WO200212272-A2.
XX
PD   14-FEB-2002.
XX
PF   03-AUG-2001; 2001WO-US24328.
XX
PR   04-AUG-2000; 2000US-223641P.
PR   13-DEC-2000; 2000US-255502P.
PR   25-JAN-2001; 2001US-264432P.
PR   26-MAR-2001; 2001US-279005P.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Nicolette CA;
XX
DR   WPI; 2002-257459/30.
XX
PT   Novel anti-melanoma peptide compounds useful for inducing immune
PT   response in a subject, and in the preparation of medicaments for the
PT   treatment and diagnosis of cancer  -
XX
PS   Claim 16; Page -; 79pp; English.
XX
CC   This invention relates to a novel anti-melanoma peptide compound
CC   comprising a peptide of the human melanoma antigen recognised by T cells
CC   (MART-1) protein. These compounds are designed to enhance binding to
CC   major histocompatibility (MHC) compounds and enhance immunoregulatory
CC   properties and induce an immune response. The invention also comprises
CC   the nucleotide sequences encoding the peptides of the invention. The
CC   compounds of the invention are useful for inducing an immune response in
CC   a subject, by delivering the compounds in the context of a major
CC   histocompatibility (MHC) molecule which presents the compound on the
CC   surface of an antigen presenting cell, or by delivering it as a
CC   polynucleotide that encodes the peptide. The invention also comprises
CC   antibodies that recognise and bind these compounds which are useful in
CC   immunotherapy. The compounds of the invention are useful for modulating
CC   an immune response to synthetic and naturally occurring compounds in a
CC   subject. The compounds may also be used as components of anti-cancer
CC   vaccines and to expand immune effector cells that are specific for
CC   cancers characterised by expression of the human melanoma antigen
CC   recognised by T cells, MART-1. The compounds of the invention are also
CC   useful for the detection and purification of antibodies and may be used
CC   for the preparation of medicaments for the diagnosis and treatment of
CC   diseases such as cancer. The compounds of the invention have enhanced
CC   binding to MHC molecules and enhanced immunoregulatory properties
CC   relative to their natural counterparts. The present sequence represents
CC   the human melanoma antigen recognised by T cells (MART-1) mutant #5 of
CC   the invention. This mutant has amino acid alterations in the region
CC   corresponding to the MHC class I binding site, these mutations confer
CC   tighter binding to the MHC.
CC   Note; This sequence is not shown in the specification but was created by
CC   the indexer from the wild type sequence shown in AA77793 and the
CC   information given in claim 16 of the specification.
XX
SQ   Sequence  118 AA;
```

```
  Query Match             99.4%;  Score 645;  DB 23;  Length 118;
  Best Local Similarity   99.2%;  Pred. No. 6e-65;
  Matches  117;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEETAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 14
AAU77809
ID   AAU77809 standard; Protein; 118 AA.
XX
AC   AAU77809;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Human MART-1 protein mutant #4.
XX
KW   MART-1; melanoma antigen recognised by T cells; human; mutant;
KW   anti-melanoma peptide; major histocompatibility complex; MHC;
KW   immunotherapy; cancer; vaccine; immunoregulatory; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 27
FT                   /note= "Wild type A substituted by F"
FT   Misc-difference 28
FT                   /note= "Wild type A substituted by A"
XX
PN   WO200212272-A2.
XX
PD   14-FEB-2002.
XX
PF   03-AUG-2001; 2001WO-US24328.
XX
PR   04-AUG-2000; 2000US-223641P.
PR   13-DEC-2000; 2000US-255502P.
PR   25-JAN-2001; 2001US-264432P.
PR   26-MAR-2001; 2001US-279005P.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Nicolette CA;
XX
DR   WPI; 2002-257459/30.
XX
PT   Novel anti-melanoma peptide compounds useful for inducing immune
PT   response in a subject, and in the preparation of medicaments for the
PT   treatment and diagnosis of cancer -
XX
PS   Claim 15; Page -; 79pp; English.
XX
CC   This invention relates to a novel anti-melanoma peptide compound
CC   comprising a peptide of the human melanoma antigen recognised by T cells
CC   (MART-1) protein. These compounds are designed to enhance binding to
CC   major histocompatibility (MHC) compounds and enhance immunoregulatory
CC   properties and induce an immune response. The invention also comprises
CC   the nucleotide sequences encoding the peptides of the invention. The
CC   compounds of the invention are useful for inducing an immune response in
CC   a subject, by delivering the compounds in the context of a major
CC   histocompatibility (MHC) molecule which presents the compound on the
CC   surface of an antigen presenting cell, or by delivering it as a
CC   polynucleotide that encodes the peptide. The invention also comprises
CC   antibodies that recognise and bind these compounds which are useful in
CC   immunotherapy. The compounds of the invention are useful for modulating
CC   an immune response to synthetic and naturally occurring compounds in a
CC   subject. The compounds may also be used as components of anti-cancer
CC   vaccines and to expand immune effector cells that are specific for
CC   cancers characterised by expression of the human melanoma antigen
CC   recognised by T cells, MART-1. The compounds of the invention are also
CC   useful for the detection and purification of antibodies and may be used
CC   for the preparation of medicaments for the diagnosis and treatment of
CC   diseases such as cancer. The compounds of the invention have enhanced
CC   binding to MHC molecules and enhanced immunoregulatory properties
CC   relative to their natural counterparts. The present sequence represents
CC   the human melanoma antigen recognised by T cells (MART-1) mutant #4 of
CC   the invention. This mutant has amino acid alterations in the region
CC   corresponding to the MHC class I binding site, these mutations confer
CC   tighter binding to the MHC.
CC   Note; This sequence is not shown in the specification but was created by
CC   the indexer from the wild type sequence shown in AA77793 and the
CC   information given in claim 15 of the specification.
XX
SQ   Sequence  118 AA;

  Query Match             99.1%;  Score 643;  DB 23;  Length 118;
  Best Local Similarity   99.2%;  Pred. No. 1e-64;
  Matches  117;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEFAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 15
AAY31980
ID   AAY31980 standard; Protein; 118 AA.
XX
AC   AAY31980;
XX
DT   21-DEC-1999  (first entry)
XX
DE   Human MART1 melanocyte differentiation antigen.
XX
KW   MART1; melanocyte differentiation antigen; melanoma; human;
KW   antigen presentation; adoptive immunotherapy; cancer; therapy;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO9947102-A2.
XX
PD   23-SEP-1999.
XX
PF   19-MAR-1999;   99WO-US06031.
XX
PR   20-MAR-1998;   98US-0078880.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Nicolette CA,  Kaplan J;
XX
DR   WPI; 1999-590956/50.
DR   N-PSDB; AAZ20065.
XX
PT   Preparing cells for use as cancer vaccines and in adoptive
PT   immunotherapy -
XX
PS   Disclosure; Page 51; 55pp; English.
XX
CC   The present sequence represents human MART1, a melanocyte
CC   differentiation antigen that is specifically recognised by HLA-A2
CC   restricted tumour-infiltrating lymphocytes of melanoma patients.
CC   The invention provides methods for immunotherapy, in particular for
CC   inducing an immune response against an antigen in a patient.
```

CC Genetically modified antigen-presenting cells (APC) that are more
CC potent presenters of exogenous peptide than the parental
CC antigen-presenting cells are used. These APCs lack an effective
CC endogenous TAP (transporter associated with antigen processing)
CC activity and present exogenous antigen on the major
CC histocompatibility complex class I (MHC-I) molecule. Suitable
CC exogenous antigens include a tumour antigen, such as a minimal
CC essential epitope of MART1, which can complex with MHC-I. The
CC genetically modified APCs are useful for inducing an immune response
CC (claimed) against an antigen in a patient (adoptive immunotherapy),
CC especially as vaccines against cancer in mammals, preferably humans.
CC The cells are also useful for expanding populations of immune
CC effector cells, preferably cytotoxic T lymphocyte cells.
XX
SQ Sequence 118 AA;

Query Match 98.9%; Score 642; DB 20; Length 118;
Best Local Similarity 99.2%; Pred. No. 1.3e-64;
Matches 117; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRPNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
```

Search completed: October 7, 2003, 18:48:16
Job time : 42 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 7, 2003, 18:43:14 ; Search time 33 Seconds
(without alignments)
922.734 Million cell updates/sec

Title: US-09-898-860-2
Perfect score: 649
Sequence: 1 MPREDAHFIYGYPKKGHGHS.....NAPPAYEKLSAEQSPPPYSP 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	420.5	64.8	113	11	Q9D1Y3	Q9d1y3 mus musculu
2	87	13.4	4957	4	O14687	O14687 homo sapien
3	87	13.4	5262	4	O14686	O14686 homo sapien
4	81	12.5	2034	2	Q93NX7	Q93nx7 streptomyce
5	80.5	12.4	344	11	Q9R201	Q9r201 mus musculu
6	80.5	12.4	1127	3	Q9P571	Q9p571 neurospora
7	79.5	12.2	494	15	Q994J1	Q994j1 human immun
8	78	12.0	540	17	Q8TUC3	Q8tuc3 methanosarc
9	78	12.0	1814	10	Q8W5N5	Q8w5n5 oryza sativ
10	77	11.9	488	4	Q8N504	Q8n504 homo sapien
11	77	11.9	491	4	Q8TBU7	Q8tbu7 homo sapien
12	77	11.9	750	10	Q9AUC2	Q9auc2 zea mays (m
13	76	11.7	303	4	Q9UKJ1	Q9ukj1 homo sapien
14	75.5	11.6	227	4	Q9UKJ0	Q9ukj0 homo sapien
15	75	11.6	107	12	Q9J455	Q9j455 human rotav
16	75	11.6	107	12	Q86198	Q86198 human rotav
17	74.5	11.5	496	15	Q8UT44	Q8ut44 human immun
18	74	11.4	739	5	Q8IGR9	Q8igr9 drosophila
19	74	11.4	883	5	Q9VBN1	Q9vbn1 drosophila
20	73.5	11.3	198	11	Q8R182	Q8r182 mus musculu
21	73	11.2	301	5	Q8MY26	Q8my26 papilio xut
22	73	11.2	1069	11	O88185	O88185 mus musculu
23	72.5	11.2	393	6	Q8MK55	Q8mk55 cyclopes di
24	72.5	11.2	405	3	Q8J221	Q8j221 glarea lozo
25	72.5	11.2	1427	5	Q9VPQ0	Q9vpq0 drosophila
26	72	11.1	1729	10	Q8RZW5	Q8rzw5 oryza sativ
27	71.5	11.0	1265	2	P72316	P72316 rhodospiril
28	71	10.9	412	4	Q9H5E0	Q9h5e0 homo sapien
29	71	10.9	500	15	Q9IVA4	Q9iva4 human immun
30	71	10.9	555	4	Q8NB59	Q8nb59 homo sapien
31	70.5	10.9	179	4	Q8TAS3	Q8tas3 homo sapien
32	70.5	10.9	254	16	Q9I0Z4	Q9i0z4 pseudomonas
33	70.5	10.9	348	2	O52641	O52641 pseudomonas
34	70.5	10.9	348	2	Q59715	Q59715 pseudomonas
35	70.5	10.9	348	2	Q51972	Q51972 pseudomonas
36	70.5	10.9	402	4	Q96L63	Q96l63 homo sapien
37	70.5	10.9	449	15	Q999X1	Q999x1 feline immu
38	70.5	10.9	449	15	Q999X2	Q999x2 feline immu
39	70.5	10.9	592	11	Q9JLN5	Q9jln5 mus musculu
40	70	10.8	112	6	Q9GMM4	Q9gmm4 macaca fasc
41	70	10.8	253	6	Q95JB6	Q95jb6 ovis aries
42	70	10.8	335	5	Q9TXV5	Q9txv5 caenorhabdi
43	70	10.8	346	10	Q9ZPT9	Q9zpt9 arabidopsis
44	70	10.8	389	16	Q8PKX9	Q8pkx9 xanthomonas
45	70	10.8	562	16	Q9K3S5	Q9k3s5 streptomyce

ALIGNMENTS

RESULT 1
Q9D1Y3
ID Q9D1Y3 PRELIMINARY; PRT; 113 AA.
AC Q9D1Y3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE A930034P04Rik protein.
GN A930034P04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK020928; BAB32259.1; -.
DR MGD; MGI:1925086; A930034P04Rik.

```
SQ   SEQUENCE   113 AA;  12818 MW;  AC01FC9840640E6E CRC64;

  Query Match             64.8%;  Score 420.5;  DB 11;  Length 113;
  Best Local Similarity   67.8%;  Pred. No. 4.3e-40;
  Matches   80;  Conservative   11;  Mismatches   22;  Indels    5;  Gaps    2;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||:|| ||  |||:||| || ||||||| ||| |:||: |||||||||||:||| ||||
Db          1 MPQEDIHF--GYPRKGHRRSYVTAEEAAGNGILIVVLGIALLIGCWYCRRRSGYRTLMDK 58

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
                 |:| |     ||  |  ||:||::| |||: :||||||||||||||    |||||||
Db         59 RRHIGIQKTSRERCSCESPDHQDSRLSSQEKSHQPVVPNAPPAYEKLS---SPPPYSP 113


RESULT 2
O14687
ID   O14687      PRELIMINARY;      PRT;  4957 AA.
AC   O14687;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ALR.
GN   ALR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97388474; PubMed=9247308;
RA   Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA   Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA   Canaani E.;
RT   "Structure and expression pattern of human ALR, a novel gene with
RT   strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT   trithorax.";
RL   Oncogene 15:549-560(1997).
CC   -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR   EMBL; AF010404; AAC51735.1; -.
DR   InterPro; IPR003889; FYrich_C.
DR   InterPro; IPR003888; FYrich_N.
DR   InterPro; IPR000910; HMG_12_box.
DR   InterPro; IPR003616; PostSET.
DR   InterPro; IPR006118; Recombinase.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 3.
DR   Pfam; PF00856; SET; 1.
DR   SMART; SM00542; FYRC; 1.
DR   SMART; SM00541; FYRN; 1.
DR   SMART; SM00398; HMG; 1.
DR   SMART; SM00249; PHD; 4.
DR   SMART; SM00508; PostSET; 1.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS00398; RECOMBINASES_2; 1.
DR   PROSITE; PS50280; SET; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   4957 AA;  531840 MW;  1026562E1419CE8D CRC64;

  Query Match             13.4%;  Score 87;  DB 4;  Length 4957;
  Best Local Similarity   31.2%;  Pred. No. 3.2;
  Matches   25;  Conservative    7;  Mismatches   30;  Indels   18;  Gaps    3;

Qy         41 LLIGCWYCRRRNGYRALMDKSLHVGTQCALTR----RCPQEGFDHRDSKVSLQEKNCEPV 96
              ||| | :| |           :| | :   |       | ||||     || |    :||
Db        939 LLIQCRHCER----------WMHAGCESLFTEDDVDHAPDEGFD----CVSCQPYVVKPV 984

Qy         97 VPNAPPAYEKLSAEQSPPPY 116
                | |||    :  ::  | |
Db        985 APVAPPELVPMKVKEPEPQY 1004


RESULT 3
O14686
ID   O14686      PRELIMINARY;      PRT;  5262 AA.
AC   O14686;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ALR.
GN   ALR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97388474; PubMed=9247308;
RA   Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA   Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA   Canaani E.;
RT   "Structure and expression pattern of human ALR, a novel gene with
RT   strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT   trithorax.";
RL   Oncogene 15:549-560(1997).
CC   -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR   EMBL; AF010403; AAC51734.1; -.
DR   Genew; HGNC:7133; MLL2.
DR   InterPro; IPR003889; FYrich_C.
DR   InterPro; IPR003888; FYrich_N.
DR   InterPro; IPR000910; HMG_12_box.
DR   InterPro; IPR003616; PostSET.
DR   InterPro; IPR006118; Recombinase.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR001965; Znf_PHD.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00628; PHD; 5.
DR   Pfam; PF00856; SET; 1.
DR   SMART; SM00542; FYRC; 1.
DR   SMART; SM00541; FYRN; 1.
DR   SMART; SM00398; HMG; 1.
DR   SMART; SM00249; PHD; 7.
DR   SMART; SM00508; PostSET; 1.
DR   SMART; SM00184; RING; 3.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS00398; RECOMBINASES_2; 1.
DR   PROSITE; PS50280; SET; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   5262 AA;  564171 MW;  26B7C74CAD417E44 CRC64;

  Query Match             13.4%;  Score 87;  DB 4;  Length 5262;
  Best Local Similarity   31.2%;  Pred. No. 3.5;
  Matches   25;  Conservative    7;  Mismatches   30;  Indels   18;  Gaps    3;

Qy         41 LLIGCWYCRRRNGYRALMDKSLHVGTQCALTR----RCPQEGFDHRDSKVSLQEKNCEPV 96
              ||| | :| |           :| | :   |       | ||||     || |    :||
Db       1244 LLIQCRHCER----------WMHAGCESLFTEDDVDHAPDEGFD----CVSCQPYVVKPV 1289

Qy         97 VPNAPPAYEKLSAEQSPPPY 116
                | |||    :  ::  | |
Db       1290 APVAPPELVPMKVKEPEPQY 1309


RESULT 4
Q93NX7
ID   Q93NX7      PRELIMINARY;      PRT;  2034 AA.
AC   Q93NX7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   AmphK.
GN   AMPHK.
OS   Streptomyces nodosus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
```

```
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
DR   EMBL; AF357202; AAK73503.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW   Oxidoreductase; Phosphopantetheine; Transferase.
SQ   SEQUENCE   2034 AA;  212063 MW;  83E962B7DC3D5747 CRC64;

  Query Match           12.5%;  Score 81;  DB 2;  Length 2034;
  Best Local Similarity  26.8%;  Pred. No. 5.6;
  Matches   33;  Conservative   14;  Mismatches   44;  Indels   32;  Gaps    6;

Qy         11 GYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSLHV------ 64
              |  |   ||   |   |   |:: ::| :           |:|   |: ::||:
Db        370 GTVKSNIGH---TQAAAGAAGVIKMVLAL----------RHG---LLPRTLHIDAPSTH 412

Qy         65 -----GTQCALTRRCP-QEGFDHRDSKVS---LQEKNCEPVVPNAPPAYEKLSAEQSPPP 115
                   |   ||   |  ||   | : ||   :   |   ::  || | |   ||| | |
Db        413 VDWDAGHVSLLTEATPWPEGEQTRRAGVSSFGISGTNAHVILEEAPAAEEDTDAEQRPEP 472

Qy        116 YSP 118
               |
Db        473 VVP 475


RESULT 5
Q9R201
ID   Q9R201      PRELIMINARY;      PRT;   344 AA.
AC   Q9R201;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   T-cell surface antigen CD2.
GN   CD2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/CJ; TISSUE=Spleen;
RA   Ma R.Z., Teuscher C.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/cByJ; TISSUE=Spleen;
RA   Ma R.Z., Cory T.;
RT   "CD2 is a candidate gene for Tmevd3 in mice.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF065909; AAD25889.1; -.
DR   EMBL; AF306543; AAG27722.1; -.
DR   HSSP; P08921; 1A64.
DR   MGD; MGI:88320; Cd2.
DR   InterPro; IPR000719; Prot_kinase.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
SQ   SEQUENCE   344 AA;  38325 MW;  3808BBF9F3E2FE3B CRC64;

  Query Match           12.4%;  Score 80.5;  DB 11;  Length 344;
  Best Local Similarity  27.4%;  Pred. No. 0.83;
  Matches   31;  Conservative   15;  Mismatches   46;  Indels   21;  Gaps    5;

Qy          3 REDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSL 62
              :|    :   |:||    | |    || |:| |:| | | | |  |:||   |   |: |
Db        188 KESKTEVVNCPEKGLSF-YVTVGVGAG-GLLLVLL-VALFIFC-ICKRRKRNRRRKDEEL 243

Qy         63 HVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPP 115
                :                 : |: |  |:  :|    |  |   :: :  |||
Db        244 EI----------------KASRTSTVERGPKPHSTPAAAAQNSVALQAPPPP 279


RESULT 6
Q9P571
ID   Q9P571      PRELIMINARY;      PRT;  1127 AA.
AC   Q9P571;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Related to actin-interacting protein AIP3.
GN   B9J10.100.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356324; CAB92017.2; -.
DR   InterPro; IPR005613; AIP3.
DR   Pfam; PF03915; AIP3; 1.
SQ   SEQUENCE   1127 AA;  122731 MW;  B6727C2CFB031DDC CRC64;

  Query Match           12.4%;  Score 80.5;  DB 3;  Length 1127;
  Best Local Similarity  25.6%;  Pred. No. 3.2;
  Matches   34;  Conservative   16;  Mismatches   52;  Indels   31;  Gaps    7;

Qy         11 GY-PKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCR------RRNGYR--ALMDKS 61
              || | :| | |      ::|:   |     :  :             ||||  :  | |
Db        223 GYGPSRGPGPS------SSGVPTRTTSTSTMASVNSGLTNMLNEGLESNGYRPDSFRDDS 276

Qy         62 LHVG-TQCALTRRCPQEGFDHRDSKVSLQE-------KNC--------EPVVPNAPPAYE 105
               : | |  : ||| | :    | |  | |         :|         |  :|  |||
Db        277 RNSGPTTASPTRRLPSQRDQSRGSVTSDQSSLSSNTMQNIPVMAPYPGEETIPTGPPAPP 336

Qy        106 KLSAEQSPPPYSP 118
              ::: :  |||  |
Db        337 EINVDNFPPPPPP 349


RESULT 7
Q994J1
ID   Q994J1      PRELIMINARY;      PRT;   494 AA.
AC   Q994J1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Gag protein (Gag polyprotein).
GN   GAG.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=98IN022;
RX   MEDLINE=21094715; PubMed=11177395;
RA   Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA   Kalish M.L., Shaw G.M., Allen S., Hahn B.H., Gao F.;
RT   "Near full-length clones and reference sequences for subtype C
RT   isolates for HIV type 1 from three different continents.";
RL   AIDS Res. Hum. Retroviruses 17:161-168(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=98IN022;
RA   Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA   Allen S., Shaw G.M., Hahn B.H., Gao F.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF286232; AAK31035.1; -.
DR   HSSP; P05888; 1AAF.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Core protein; Polyprotein.
SQ   SEQUENCE   494 AA;  55092 MW;  A2C66A4EA3DB5E35 CRC64;

  Query Match             12.2%;  Score 79.5;  DB 15;  Length 494;
  Best Local Similarity   21.7%;  Pred. No. 1.6;
  Matches   28;  Conservative   13;  Mismatches   49;  Indels   39;  Gaps    4;

Qy         16 GHGHSYTTAEEAAGIGILTVIL------GVLLLIGCWYCRRRNGYRALMDKSLHVGTQCA 69
              | ||      ||      |:::      |   :: |: |           |  |:   |
Db        352 GPGHKARVLAEAMSQANSTIMMQRGNFKGPKRIVKCFNC----------GKEGHIAKNCR 401

Qy         70 LTR-----RCPQEGF-----------------DHRDSKVSLQEKNCEPVVPNAPPAYEK 106
                 |     :| :||                     |:    :  :   || || ||||
Db        402 APRKKGCWKCGKEGHQMKDCTERQANFLGRIWPSHKGRPGNFLQSRPEPTVPTAPPAESF 461

Qy        107 LSAEQSPPP 115
                | :| |
Db        462 GFGETTPAP 470


RESULT 8
Q8TUC3
ID   Q8TUC3      PRELIMINARY;      PRT;   540 AA.
AC   Q8TUC3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Methylamine methyltransferase corrinoid activation protein.
GN   RAMA OR MA0150.
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238;
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
DR   EMBL; AE010672; AAM03603.1; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   Pfam; PF00037; fer4; 2.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW   Transferase; Methyltransferase; Complete proteome.
SQ   SEQUENCE   540 AA;  58864 MW;  7B3D323035D5227B CRC64;

  Query Match             12.0%;  Score 78;  DB 17;  Length 540;
  Best Local Similarity   26.3%;  Pred. No. 2.7;
  Matches   30;  Conservative   17;  Mismatches   31;  Indels   36;  Gaps    8;

Qy          9 IYGYPKKGHGHSYTTAE-------EAAGIGILTVI--LGVLLLIG---CWYCRR------ 50
              :|| ||      :   |       |  |:| | ::  :|::| :    | ||::
Db        432 LYGLPKLPRTLEHADIEKRVSKDIEQVGVGGLAILKEIGIILEVSVDKCIYCQKCVEECP 491

Qy         51 --------RNGYR-ALMDKSLHVGTQCALTRR----CPQEGFDHRDSKVSLQEK 91
                      || | |  |    :|| |   ||    ||:   |   :|: ::||
Db        492 EGALEIVETNGNRIAKYDSQKCLGTSC---RRCVSICPENALD--ITKLKIKEK 540


RESULT 9
Q8W5N5
ID   Q8W5N5      PRELIMINARY;      PRT;  1814 AA.
AC   Q8W5N5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative disease resistant protein.
GN   OSJNBA0001K12.15.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA   Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA   Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA   VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA   Salzberg S.L., White O., Fraser C.M.;
RT   "Oryza sativa chromosome 10 BAC OSJNBa0001K12 genomic sequence.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC078948; AAL31031.1; -.
DR   Gramene; Q8W5N5; -.
DR   InterPro; IPR000767; Disease_resist.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR002182; NB-ARC.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF00931; NB-ARC; 1.
DR   PRINTS; PR00364; DISEASERSIST.
SQ   SEQUENCE   1814 AA;  199893 MW;  C18A0AEDE6C4DD27 CRC64;

  Query Match             12.0%;  Score 78;  DB 10;  Length 1814;
  Best Local Similarity   27.5%;  Pred. No. 11;
  Matches   28;  Conservative   11;  Mismatches   33;  Indels   30;  Gaps    5;

Qy         30 IGILTVILGVLLLIGCWYCRRRNGYRALMDKSLHVGTQCALTRRCPQ---------- 76
              : || :|    ||   |       || ||::  || :       |||
Db       1599 VAILLIISVAQLLFTFWM------YRKLMNQNSKVGKEATSNSDCPQFFSPSFSLLTFAL 1652

Qy         77 EGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              | |       |: ::  :|: |  ||  ||   |   ||  |
Db       1653 ESF-------SVVDEEVKPISP--PP--EKEEEEAISPPQPP 1683


RESULT 10
```

```
Q8N504
ID   Q8N504      PRELIMINARY;      PRT;   488 AA.
AC   Q8N504;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033132; AAH33132.1; -.
DR   InterPro; IPR006586; ADAM_cysteine.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00200; disintegrin; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; Disintegrin; 1.
DR   SMART; SM00608; ACR; 1.
DR   SMART; SM00050; DISIN; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   488 AA;  52722 MW;  F9D288B23529BC05 CRC64;

  Query Match             11.9%;  Score 77;  DB 4;  Length 488;
  Best Local Similarity   25.5%;  Pred. No. 3.1;
  Matches   35;  Conservative   16;  Mismatches   54;  Indels   32;  Gaps    7;

Qy          4 EDAHFIYGYP-----KKGHGHSYTTA---EEAAG---IGILTVILGVLLLIGCWYCRRRN 52
              :: | : |:        ||| |  :     |: |    |:|  || : :|:  :|| |:|
Db        238 QNCHCLPGWAPPFCNTPGHGGSIDSGPMPPESVGPVVAGVLVAILVLAVLMLMYYCCRQN 297

Qy         53 GYRALMDKSLHVGTQCALTRR------CP-----QEGFDHRDSKVSLQEKNCEPVVPNAP 101
                :  |       ||  :      ||       |  | :    ||    :  | | |
Db        298 NKLGQLKPS-------ALPSKLRQQFSCPFRVSQNSGTGHANPTFKLQTPQGKRKVINTP 350

Qy        102 PAYEKLSAEQSPPPYSP 118
              | |     |||  |
Db        351 ---EILRKPSQPPPRPP 364


RESULT 11
Q8TBU7
ID   Q8TBU7      PRELIMINARY;      PRT;   491 AA.
AC   Q8TBU7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Strausberg R.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC024214; AAH24214.1; -.
DR   InterPro; IPR006586; ADAM_cysteine.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00200; disintegrin; 1.
DR   ProDom; PD000664; Disintegrin; 1.
DR   SMART; SM00608; ACR; 1.
DR   SMART; SM00050; DISIN; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   491 AA;  53045 MW;  BACD338C38496A08 CRC64;

  Query Match             11.9%;  Score 77;  DB 4;  Length 491;
  Best Local Similarity   25.5%;  Pred. No. 3.1;
  Matches   35;  Conservative   16;  Mismatches   54;  Indels   32;  Gaps    7;

Qy          4 EDAHFIYGYP-----KKGHGHSYTTA---EEAAG---IGILTVILGVLLLIGCWYCRRRN 52
              :: | : |:        ||| |  :     |: |    |:|  || : :|:  :|| |:|
Db        241 QNCHCLPGWAPPFCNTPGHGGSIDSGPMPPESVGPVVAGVLVAILVLAVLMLMYYCCRQN 300

Qy         53 GYRALMDKSLHVGTQCALTRR------CP-----QEGFDHRDSKVSLQEKNCEPVVPNAP 101
                :  |       ||  :      ||       |  | :    ||    :  | | |
Db        301 NKLGQLKPS-------ALPSKLRQQFSCPFRVSQNSGTGHANPTFKLQTPQGKRKVINTP 353

Qy        102 PAYEKLSAEQSPPPYSP 118
              | |     |||  |
Db        354 ---EILRKPSQPPPRPP 367


RESULT 12
Q9AUC2
ID   Q9AUC2      PRELIMINARY;      PRT;   750 AA.
AC   Q9AUC2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Receptor-like protein kinase 1.
GN   PRK1.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. B73;
RA   Kim H.-U., Cotter R., McCormick S.;
RT   "New pollen-expressed receptor kinases identified in tomato, maize and
RT   Arabidopsis: the tomato kinases show distinct but overlapping
RT   expression patterns during pollen tube growth.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. B73;
RA   Dodds P., Kulikauskas R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF243041; AAK28346.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007090; LRR_plant.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF00560; LRR; 4.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS50502; LRR_PS; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Kinase; Transferase.
SQ   SEQUENCE   750 AA;  79386 MW;  34B12797F000351B CRC64;

  Query Match             11.9%;  Score 77;  DB 10;  Length 750;
  Best Local Similarity   24.3%;  Pred. No. 5.1;
  Matches   25;  Conservative   14;  Mismatches   40;  Indels   24;  Gaps    3;

Qy         14 KKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWY--CRRRNGYRALMDKSLHVGTQCALT 71
              |:  ||       :   |:|   :| | : |  :   ||| ||     |: : |    :
Db        305 KQDEGHKPVEGSTSTSFGVLAAFIGTLGIAGVAFVALRRRRGY-----KTKNFGPTAS-- 357

Qy         72 RRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPP 114
```

```
                          | :  :   | |: | |  : || |:||
Db        358 --------------SARPSDPPRVEPHPPAAKAEASAAQAPP 385
```

```
RESULT 13
Q9UKJ1
ID   Q9UKJ1      PRELIMINARY;      PRT;   303 AA.
AC   Q9UKJ1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Inhibitory receptor PILRalpha.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20127940; PubMed=10660620;
RA   Mousseau D.D., Banville D., L'Abbe D., Bouchard P., Shen S.H.;
RT   "PILRalpha, a novel immunoreceptor tyrosine-based inhibitory motif-
RT   bearing protein, recruits SHP-1 upon tyrosine phosphorylation and is
RT   paired with the truncated counterpart PILRbeta.";
RL   J. Biol. Chem. 275:4467-4474(2000).
DR   EMBL; AF161080; AAD52964.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   SMART; SM00409; IG; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Receptor.
SQ   SEQUENCE   303 AA;  33877 MW;  0410ADFC7E80928B CRC64;

  Query Match             11.7%;  Score 76;  DB 4;  Length 303;
  Best Local Similarity   20.8%;  Pred. No. 2.3;
  Matches   25;  Conservative   22;  Mismatches   39;  Indels   34;  Gaps    5;

Qy         21 YTTAEEAAGIGILTVILGVLLL-IGC---WYCRRRNGYRALMDKSLHVGTQCALTRRCPQ 76
              : : | | |: :   :||:::| : |   |  ||| | :          |:      | |
Db        191 HISLETAVGVAVAVTVLGIMILGLICLLRW--RRRKGQQR---------TKATTPAREPF 239

Qy         77 EGFDHRDSKVSLQEKNCEPVV------------------PNAPPAYEKLSAEQSPPPYS 117
              :  :     :  : :| :| :                  | |||::  | : |:   ||
Db        240 QNTEEPYENIRNEGQNTDPKLNPKDDGIVYASLALSSSTSPKAPPSHRPLKSPQNETLYS 299


RESULT 14
Q9UKJ0
ID   Q9UKJ0      PRELIMINARY;      PRT;   227 AA.
AC   Q9UKJ0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Activating receptor PILRbeta.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20127940; PubMed=10660620;
RA   Mousseau D.D., Banville D., L'Abbe D., Bouchard P., Shen S.H.;
RT   "PILRalpha, a novel immunoreceptor tyrosine-based inhibitory motif-
RT   bearing protein, recruits SHP-1 upon tyrosine phosphorylation and is
RT   paired with the truncated counterpart PILRbeta.";
RL   J. Biol. Chem. 275:4467-4474(2000).
DR   EMBL; AF161081; AAD52965.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   SMART; SM00409; IG; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Receptor.
SQ   SEQUENCE   227 AA;  25542 MW;  7FF960C60AB7EF09 CRC64;

  Query Match             11.6%;  Score 75.5;  DB 4;  Length 227;
  Best Local Similarity   43.8%;  Pred. No. 1.9;
  Matches   21;  Conservative    6;  Mismatches   14;  Indels    7;  Gaps    3;

Qy         15 KGHGHSY-----TTAEEAAGIGIL-TVILGVLLLIGCWYCRRRNGYRA 56
              |||  |:     |    |  : :| |||||:| |:  |: ||| | ||
Db        176 KGHSESWHLSLDTAIRVALAVAVLKTVILGLLCLLLLWW-RRRKGSRA 222


RESULT 15
Q9J455
ID   Q9J455      PRELIMINARY;      PRT;   107 AA.
AC   Q9J455;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Putative fusion protein (Nonstructural protein).
OS   Human rotavirus B.
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=10942;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CAL-1;
RA   Sen A., Das S., Krishnan T., Kobayashi N., Naik T.N.;
RT   "partial sequence of cDNA of the gene segment 6 encoding fusion
RT   protein and a nonstructural protein of the human group B rotavirus
RT   strain, CAL-1.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CAL;
RA   Kobayashi N., Naik T.;
RT   "Sequence of the sixth RNA segment of group B human rotavirus CAL.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF203912; AAF69263.1; -.
DR   EMBL; AF230975; AAG09741.1; -.
SQ   SEQUENCE   107 AA;  11697 MW;  3AD1596D92B387FE CRC64;

  Query Match             11.6%;  Score 75;  DB 12;  Length 107;
  Best Local Similarity   38.1%;  Pred. No. 0.92;
  Matches   16;  Conservative   12;  Mismatches    8;  Indels    6;  Gaps    2;

Qy         27 AAGIGILTVILGVLL---LIGCWYCR---RRNGYRALMDKSL 62
              |||:||:  :| :||   :: |: ||   | ||  :|::::|
Db         42 AAGVGIIATLLVLLLCSCVLNCYLCRKLKRTNGVSSLLERNL 83


Search completed: October  7, 2003, 18:49:15
Job time : 35 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 7, 2003, 18:42:34 ; Search time 11 Seconds
(without alignments)
504.468 Million cell updates/sec

Title: US-09-898-860-2
Perfect score: 649
Sequence: 1 MPREDAHFIYGYPKKGHGHS.....NAPPAYEKLSAEQSPPPYSP 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	649	100.0	118	1	MAR1_HUMAN	Q16655 homo sapien
2	80.5	12.4	344	1	CD2_MOUSE	P08920 mus musculu
3	77	11.9	710	1	NASC_BACSU	P42434 bacillus su
4	77	11.9	956	1	AD19_HUMAN	Q9h013 homo sapien
5	72.5	11.2	815	1	AD15_MOUSE	O88839 mus musculu
6	72.5	11.2	1429	1	EXPA_DROME	Q07436 drosophila
7	71	10.9	1069	1	PCH7_HUMAN	O60245 homo sapien
8	70.5	10.9	202	1	TMG2_HUMAN	O14669 homo sapien
9	69.5	10.7	218	1	TMG1_HUMAN	O14668 homo sapien
10	69.5	10.7	401	1	LSG1_HAEIN	P71399 haemophilus
11	68	10.5	165	1	CYY1_MOUSE	Q8vih7 mus musculu
12	67.5	10.4	188	1	COME_METJA	P58416 methanococc
13	67.5	10.4	778	1	MGD1_HUMAN	Q9y5v3 homo sapien
14	67.5	10.4	4590	1	FATH_HUMAN	Q14517 homo sapien
15	66.5	10.2	775	1	MGD1_MOUSE	Q9qyh6 mus musculu
16	66.5	10.2	775	1	MGD1_RAT	Q9es73 rattus norv
17	66	10.2	140	1	YD23_YEAST	Q07549 saccharomyc
18	66	10.2	969	1	SACB_STRSL	Q55242 streptococc
19	65.5	10.1	258	1	BOX5_NOTVI	P53771 notophthalm
20	65.5	10.1	659	1	SYT_THETH	P56881 thermus the
21	65.5	10.1	1436	1	WC11_BOVIN	P30205 bos taurus
22	65	10.0	309	1	YHCC_ECOLI	P45476 escherichia
23	65	10.0	356	1	RS41_ARATH	P92966 arabidopsis
24	64.5	9.9	154	1	CYY1_HUMAN	Q96j86 homo sapien
25	64.5	9.9	359	1	FIXZ_RHILE	P07748 rhizobium l
26	64	9.9	194	1	TIM2_RABIT	Q9trz7 oryctolagus
27	64	9.9	344	1	CD2_RAT	P08921 rattus norv
28	64	9.9	352	1	CD5L_MOUSE	Q9qwk4 mus musculu
29	64	9.9	2180	1	POLG_EC22H	Q66578 e genome po
30	63.5	9.8	118	1	VPX_SIVAI	Q02842 simian immu
31	63.5	9.8	227	1	NK11_MOUSE	P27811 mus musculu
32	63.5	9.8	628	1	GIDA_WIGBR	Q8d3k0 wiggleswort
33	63.5	9.8	1338	1	ACIN_MOUSE	Q9jix8 mus musculu
34	63	9.7	117	1	HYPA_HELPY	O25539 helicobacte
35	63	9.7	336	1	EFB2_MOUSE	P52800 mus musculu
36	63	9.7	387	1	GT4R_HUMAN	Q9nr83 homo sapien
37	63	9.7	629	1	GIDA_HAEIN	P44763 haemophilus
38	63	9.7	1004	1	PHC1_HUMAN	P78364 homo sapien
39	63	9.7	1012	1	PHC1_MOUSE	Q64028 mus musculu
40	62.5	9.6	757	1	KNC3_HUMAN	Q14003 homo sapien
41	62.5	9.6	786	1	CT32_HUMAN	Q9nq75 homo sapien
42	62.5	9.6	821	1	ENV_SIVGB	P22380 simian immu
43	62.5	9.6	889	1	KNC3_RAT	Q01956 rattus norv
44	62.5	9.6	913	1	DDR1_HUMAN	Q08345 h epithelia
45	62	9.6	286	1	TESB_HAEIN	P44498 haemophilus

ALIGNMENTS

RESULT 1
MAR1_HUMAN
ID MAR1_HUMAN STANDARD; PRT; 118 AA.
AC Q16655;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Melanoma antigen recognized by T-cells 1 (MART-1) (Melan-A protein)
DE (Antigen SK29-AA) (Antigen LB39-AA).
GN MLANA OR MART1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=94224770; PubMed=8170938;
RA Kawakami Y., Eliyahu S., Delgado C.H., Robbins P.F., Rivoltini L.,
RA Topalian S.L., Miki T., Rosenberg S.A.;
RT "Cloning of the gene coding for a shared human melanoma antigen
RT recognized by autologous T cells infiltrating into tumor.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3515-3519(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94275389; PubMed=8006593;
RA Coulie P.G., Brichard V., van Pel A., Woelfel T., Schneider J.,
RA Traversari C., Mattei S., de Plaen E., Lurquin C., Szikora J.-P.,
RA Renauld J.-C., Boon T.;
RT "A new gene coding for a differentiation antigen recognized by
RT autologous cytolytic T lymphocytes on HLA-A2 melanomas.";
RL J. Exp. Med. 180:35-42(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

```
CC   -!- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO MELANOMA AND
CC       MELANOCYTE CELL LINES AND RETINA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U06452; AAA19238.1; -.
DR   EMBL; U06654; AAA20389.1; -.
DR   EMBL; BC014423; AAH14423.1; -.
DR   PIR; A55253; A55253.
DR   Genew; HGNC:7124; MLANA.
DR   MIM; 605513; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0008222; F:tumor antigen; TAS.
KW   Antigen; Transmembrane.
FT   TRANSMEM     27     47       POTENTIAL.
SQ   SEQUENCE   118 AA;  13157 MW;  B755BFF39CFCB16E CRC64;

  Query Match             100.0%;  Score 649;  DB 1;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-61;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Qy         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT 2
CD2_MOUSE
ID   CD2_MOUSE      STANDARD;      PRT;   344 AA.
AC   P08920; Q61394;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   T-cell surface antigen CD2 precursor (T-cell surface antigen
DE   T11/Leu-5) (LFA-2) (LFA-3 receptor).
GN   CD2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10.A;
RX   MEDLINE=87276135; PubMed=2440689;
RA   Sewell W.A., Brown M.H., Fink P.J., Kozak C.A., Crumpton M.J.;
RT   "The murine homologue of the T lymphocyte CD2 antigen: molecular
RT   cloning, chromosome assignment and cell surface expression.";
RL   Eur. J. Immunol. 17:1015-1020(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88004738; PubMed=2820751;
RA   Clayton L.K., Sayre P.H., Novotny J., Reinherz E.L.;
RT   "Murine and human T11 (CD2) cDNA sequences suggest a common signal
RT   transduction mechanism.";
RL   Eur. J. Immunol. 17:1367-1370(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=88144486; PubMed=2894031;
RA   Diamond D.J., Clayton L.K., Sayre P.H., Reinherz E.L.;
RT   "Exon-intron organization and sequence comparison of human and murine
RT   T11 (CD2) genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:1615-1619(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88140313; PubMed=3257775;
RA   Yagita H., Okumura K., Nakauchi H.;
RT   "Molecular cloning of the murine homologue of CD2. Homology of the
RT   molecule to its human counterpart T11.";
RL   J. Immunol. 140:1321-1326(1988).
RN   [5]
RP   INTERACTION WITH CD2AP.
RX   MEDLINE=98412662; PubMed=9741631;
RA   Dustin M.L., Olszowy M.W., Holdorf A.D., Li J., Bromley S., Desai N.,
RA   Widder P., Rosenberger F., van der Merwe P.A., Allen P.M., Shaw A.S.;
RT   "A novel adaptor protein orchestrates receptor patterning and
RT   cytoskeletal polarity in T-cell contacts.";
RL   Cell 94:667-677(1998).
CC   -!- FUNCTION: CD2 interacts with lymphocyte function-associated
CC       antigen (LFA-3) and CD48/BCM1 to mediate adhesion between T
CC       cells and other cell types. CD2 is implicated in the triggerring
CC       of T-cells, the cytoplasmic domain is implicated in the
CC       signaling function.
CC   -!- SUBUNIT: Interacts with CD2AP.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y00023; CAA68258.1; -.
DR   EMBL; X06143; CAA29500.1; -.
DR   EMBL; M19807; AAA37393.1; -.
DR   EMBL; M19799; AAA37393.1; JOINED.
DR   EMBL; M19801; AAA37393.1; JOINED.
DR   EMBL; M19803; AAA37393.1; JOINED.
DR   EMBL; M19805; AAA37393.1; JOINED.
DR   EMBL; M18934; AAA37397.1; -.
DR   PIR; I49585; I49585.
DR   HSSP; P08921; 1A64.
DR   MGD; MGI:88320; Cd2.
DR   GO; GO:0005515; F:protein binding activity; ISS.
DR   GO; GO:0004872; F:receptor activity; ISS.
DR   PROSITE; PS50835; IG_LIKE; FALSE_NEG.
KW   Immunoglobulin domain; T-cell; Glycoprotein; Antigen; Transmembrane;
KW   Cell adhesion; Repeat; Signal.
FT   SIGNAL       1     22
FT   CHAIN       23    344       T-CELL SURFACE ANTIGEN CD2.
FT   DOMAIN      23    203       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    204   229       POTENTIAL.
FT   DOMAIN      230   344       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      23    121       IG-LIKE V-TYPE.
FT   DOMAIN      122   202       IG-LIKE C2-TYPE.
FT   DOMAIN      276   343       PRO-RICH.
FT   DISULFID    133   197       BY SIMILARITY.
FT   DISULFID    140   180       BY SIMILARITY.
FT   CARBOHYD    82    82        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    94    94        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    135   135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    166   166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    99    99        Y -> T (IN REF. 3).
FT   CONFLICT    128   128       M -> V (IN REF. 3 AND 4).
FT   CONFLICT    139   139       T -> I (IN REF. 4).
FT   CONFLICT    175   175       N -> A (IN REF. 3).
FT   CONFLICT    175   175       N -> S (IN REF. 4).
FT   CONFLICT    191   191       K -> N (IN REF. 2).
FT   CONFLICT    192   192       M -> T (IN REF. 3 AND 4).
SQ   SEQUENCE   344 AA;  38414 MW;  CFD12FCBD1444450 CRC64;

  Query Match             12.4%;  Score 80.5;  DB 1;  Length 344;
```

```
  Best Local Similarity   27.4%;  Pred. No. 0.44;
  Matches   31;  Conservative   15;  Mismatches   46;  Indels   21;  Gaps    5;

Qy          3 REDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSL 62
              :|    :   |:||    | |    || |:| |:| | | | |  |:||   |   |: |
Db        188 KESKMEVVNCPEKGLSF-YVTVGVGAG-GLLLVLL-VALFIFC-ICKRRKRNRRRKDEEL 243

Qy         63 HVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPP 115
                :                : |: |  |:  :|    |  |   :: :  |||
Db        244 EI----------------KASRTSTVERGPKPHSTPAAAAQNSVALQAPPPP 279


RESULT 3
NASC_BACSU
ID   NASC_BACSU     STANDARD;      PRT;   710 AA.
AC   P42434;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Assimilatory nitrate reductase catalytic subunit (EC 1.7.99.4).
GN   NASC OR NARB OR NASBB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124189; PubMed=8969502;
RA   Yamane K., Kumano M., Kurita K.;
RT   "The 25 degrees-36 degrees region of the Bacillus subtilis
RT   chromosome: determination of the sequence of a 146 kb segment and
RT   identification of 113 genes.";
RL   Microbiology 142:3047-3056(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S.,'Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 35-710 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95173124; PubMed=7868621;
RA   Ogawa K.-I., Akagawa E., Yamane K., Sun Z.-W., Lacelle M., Zuber P.,
RA   Nakano M.M.;
RT   "The nasB operon and nasA gene are required for nitrate/nitrite
RT   assimilation in Bacillus subtilis.";
RL   J. Bacteriol. 177:1409-1413(1995).
CC   -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC       STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC   -!- CATALYTIC ACTIVITY: Nitrite + acceptor = nitrate + reduced
CC       acceptor.
CC   -!- COFACTOR: MOLYBDENUM (MOLYBDOPTERIN); MAY BIND A 4FE-4S CLUSTER.
CC   -!- PATHWAY: Nitrate assimilation (denitrification).
CC   -!- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTERIN-CONTAINING
CC       OXIDOREDUCTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D50453; BAA08965.1; -.
DR   EMBL; Z99105; CAB12125.1; -.
DR   EMBL; D30689; BAA06353.1; -.
DR   PIR; E69665; E69665.
DR   HSSP; P81186; 2NAP.
DR   SubtiList; BG11095; nasC.
DR   InterPro; IPR006657; Mol_dinuc_bind.
DR   InterPro; IPR006963; Molybdop_Fe4S4.
DR   InterPro; IPR006656; Molybdopterin.
DR   InterPro; IPR006655; Prok_Mboxred.
DR   Pfam; PF04879; Molybdop_Fe4S4; 1.
DR   Pfam; PF00384; molybdopterin; 1.
DR   Pfam; PF01568; Molydop_binding; 1.
DR   PROSITE; PS00551; MOLYBDOPTERIN_PROK_1; 1.
DR   PROSITE; PS00490; MOLYBDOPTERIN_PROK_2; 1.
DR   PROSITE; PS00932; MOLYBDOPTERIN_PROK_3; FALSE_NEG.
KW   Oxidoreductase; Molybdenum; Nitrate assimilation; Iron-sulfur; 4Fe-4S;
KW   Complete proteome.
FT   METAL        26     26       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL        29     29       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL        33     33       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL        63     63       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   710 AA;  78575 MW;  625E8864A1552AA2 CRC64;

  Query Match             11.9%;  Score 77;  DB 1;  Length 710;
  Best Local Similarity   32.4%;  Pred. No. 2.2;
  Matches   34;  Conservative    8;  Mismatches   37;  Indels   26;  Gaps    6;

Qy         15 KGHGHSYTTAEE------AAGIGILTVILGVLLLIGCWY--CRRRNGYRALMDKSLHVGT 66
              ||   |||:||:       |  |      |:    |  |   ||  |      :| | ||
Db        507 KGRYFSYTSAEDIFNELREASRG------GIADYSGISYGRLRREGGIHWPCPESDHPGT 560

Qy         67 QCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPA-YEKLSAE 110
                |    | | | | | :|       |:|| ||   || :|:
Db        561 GRLFT-----ESFAHPDQKAALS------VIPNEPPVPKEKPTAD 594


RESULT 4
AD19_HUMAN
ID   AD19_HUMAN     STANDARD;      PRT;   956 AA.
AC   Q9H013; Q9BZL5; Q9UHP2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   ADAM 19 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE   domain 19) (Meltrin beta) (Metalloprotease and disintegrin dentritic
DE   antigen marker) (MADDAM).
GN   ADAM19 OR MLTNB OR FKSG34.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM A).
RA   Wang Y.-G., Gong L.;
RT   "Identification of FKSG34, a novel human gene encoding for
RT   metalloprotease-disintegrin meltrin beta.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM B).
RC   TISSUE=Lymph node;
RX   MEDLINE=20346928; PubMed=10887142;
RA   Fritsche J., Moser M., Faust S., Peuker A., Buettner R., Andreesen R.,
RA   Kreutz M.;
RT   "Molecular cloning and characterization of a human metalloprotease
RT   disintegrin a novel marker for dendritic cell differentiation.";
RL   Blood 96:732-739(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM B).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=21092672; PubMed=11162584;
RA   Wei P., Zhao Y.-G., Zhuang L., Ruben S., Sang Q.-X.A.;
RT   "Expression and enzymatic activity of human disintegrin and
RT   metalloproteinase ADAM19/meltrin beta.";
RL   Biochem. Biophys. Res. Commun. 280:744-755(2001).
RN   [4]
RP   SEQUENCE OF 100-956 FROM N.A. (ISOFORM A).
RA   Xu R., Cai J., Ying B., Wang F., Xu T., Zhao S., Li C.;
RT   "Partial sequence of Homo sapiens ADAM19.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PARTICIPATES IN THE PROTEOLYTIC PROCESSING OF BETA-TYPE
CC       NEUREGULIN ISOFORMS WHICH ARE INVOLVED IN NEUROGENESIS AND
CC       SYNAPTOGENESIS, SUGGESTING A REGULATORY ROLE IN GLIAL CELL. ALSO
CC       CLEAVES ALPHA-2 MACROGLOBULIN. MAY BE INVOLVED IN OSTEOBLAST
CC       DIFFERENCIATION AND/OR OSTEOBLAST ACTIVITY IN BONE (BY
CC       SIMILARITY).
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=A;
CC         IsoId=Q9H013-1; Sequence=Displayed;
CC       Name=B;
CC         IsoId=Q9H013-2; Sequence=VSP_005481;
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MANY NORMAL ORGANS TISSUES AND
CC       SEVERAL CANCER CELL LINES.
CC   -!- INDUCTION: BY 1,25(OH)2VD3 IN MONOCYTES.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: Belongs to peptidase family M12B.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 disintegrin domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF326918; AAG50282.1; -.
DR   EMBL; Y13786; CAC20585.1; -.
DR   EMBL; AF311317; AAK07852.1; -.
DR   EMBL; AF134707; AAF22162.1; -.
DR   HSSP; P18619; 1FVL.
DR   Genew; HGNC:197; ADAM19.
DR   MIM; 603640; -.
DR   InterPro; IPR006586; ADAM_cysteine.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001818; Matrixin.
DR   InterPro; IPR002870; Pep_M12B_propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; Disintegrin; 1.
DR   SMART; SM00608; ACR; 1.
DR   SMART; SM00050; DISIN; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00546; CYSTEINE_SWITCH; FALSE_NEG.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Transmembrane; EGF-like domain; SH3-binding; Alternative splicing.
FT   SIGNAL       1     25       POTENTIAL.
FT   PROPEP      26    203       BY SIMILARITY.
FT   CHAIN      204    956       ADAM 19.
FT   DOMAIN     204    700       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   701    721       POTENTIAL.
FT   DOMAIN     722    956       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     204    410       METALLOPROTEASE.
FT   DOMAIN     416    502       DISINTEGRIN-LIKE.
FT   DOMAIN     435    438       POLY-GLU.
FT   DOMAIN     503    650       CYS-RICH.
FT   DOMAIN     651    683       EGF-LIKE.
FT   SITE       834    840       SH3-BINDING (POTENTIAL).
FT   SITE       839    845       SH3-BINDING (POTENTIAL).
FT   SITE       133    133       CYSTEINE SWITCH (POTENTIAL).
FT   METAL      346    346       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   347    347       BY SIMILARITY.
FT   METAL      350    350       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      356    356       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID   321    404       BY SIMILARITY.
FT   DISULFID   361    388       BY SIMILARITY.
FT   DISULFID   475    482       POTENTIAL.
FT   DISULFID   655    665       BY SIMILARITY.
FT   DISULFID   659    671       BY SIMILARITY.
FT   DISULFID   673    682       BY SIMILARITY.
FT   CARBOHYD   145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   445    445       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   646    646       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   903    956       VSPREALKVKAGTRGLQGGRCRVEKTKQFMLLVVWTELPEQ
FT                               KPRAKHSCFLVPA -> FPEYRSQRAGGMISSKI (in
FT                               isoform B).
FT                               /FTId=VSP_005481.
FT   CONFLICT     4      4       G -> S (IN REF. 2).
FT   CONFLICT    32     33       SK -> R (IN REF. 2 AND 3).
FT   CONFLICT   558    558       V -> D (IN REF. 2 AND 3).
FT   CONFLICT   623    623       N -> D (IN REF. 2 AND 3).
SQ   SEQUENCE   956 AA;  105038 MW;  8373F10FA0418B12 CRC64;

  Query Match             11.9%;  Score 77;  DB 1;  Length 956;
  Best Local Similarity   25.5%;  Pred. No. 3.1;
  Matches   35;  Conservative   16;  Mismatches   54;  Indels   32;  Gaps    7;

Qy          4 EDAHFIYGYP-----KKGHGHSYTTA---EEAAG---IGILTVILGVLLLIGCWYCRRRN 52
              :: | : |:         ||| | :     |: |    |:|  || : :|:  :|| |:|
Db        669 QNCHCLPGWAPPFCNTPGHGGSIDSGPMPPESVGPVVAGVLVAILVLAVLMLMYYCCRQN 728

Qy         53 GYRALMDKSLHVGTQCALTRR------CP-----QEGFDHRDSKVSLQEKNCEPVVPNAP 101
                :  |       ||  :      ||       |  | :    ||    :  | | |
Db        729 NKLGQLKPS-------ALPSKLRQQFSCPFRVSQNSGTGHANPTFKLQTPQGKRKVINTP 781

Qy        102 PAYEKLSAEQSPPPYSP 118
                | |     |||  |
Db        782 ---EILRKPSQPPPRPP 795


RESULT 5
AD15_MOUSE
```

```
ID   AD15_MOUSE     STANDARD;      PRT;   815 AA.
AC   O88839; Q9QYL2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   ADAM 15 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE   domain 15) (Metalloproteinase-like, disintegrin-like, and cysteine-
DE   rich protein 15) (MDC-15) (Metalloprotease RGD disintegrin protein)
DE   (Metargidin) (AD56).
GN   ADAM15 OR MDC15.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND PROCESSING.
RC   TISSUE=Lung;
RX   MEDLINE=98421554; PubMed=9748307;
RA   Lum L., Reid M.S., Blobel C.P.;
RT   "Intracellular maturation of the mouse metalloprotease disintegrin
RT   MDC15.";
RL   J. Biol. Chem. 273:26236-26247(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Myeloid, and Myeloma;
RA   Shimizu E., Higuchi Y., Matsuura K., Hijiya N., Yamamoto S.;
RT   "Structure of the mouse ADAM 15 (AD56) gene.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   INTERACTION WITH ENDOPHILIN I AND SNX9.
RX   MEDLINE=20002705; PubMed=10531379;
RA   Howard L., Nelson K.K., Maciewicz R.A., Blobel C.P.;
RT   "Interaction of the metalloprotease disintegrins MDC9 and MDC15 with
RT   two SH3 domain-containing proteins, endophilin I and SH3PX1.";
RL   J. Biol. Chem. 274:31693-31699(1999).
CC   -!- FUNCTION: May be involved in cell-surface proteolysis, cell
CC       adhesion or intracellular protein maturation.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (Potential).
CC   -!- SUBUNIT: Interacts specifically with Src family protein-tyrosine
CC       kinases (PTKs) (By similarity). Interacts with ITAGV-ITGB3
CC       (Vitronectin receptor), PACSIN3 and SNX9. PACSIN3 and SNX9
CC       preferentially bind the precursor but not the processed form of
CC       ADAM15, suggesting that the interaction occurs in a secretory
CC       pathway compartment prior to the medial Golgi.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THE MAJORITY OF THE
CC       PROTEIN IS LOCALIZED IN A PERINUCLEAR COMPARTMENT WHICH MAY
CC       CORRESPOND TO THE TRANS-GOLGI NETWORK OR THE LATE ENDOSOME. THE
CC       PRO-PROTEIN IS THE MAJOR DETECTABLE FORM ON THE CELL SURFACE,
CC       WHEREAS THE MAJORITY OF THE PROTEIN IN THE CELL IS PROCESSED.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, BRAIN, LUNG, AND
CC       KIDNEY. EXPRESSED AT LOWER LEVELS IN SPLEEN, LIVER, TESTIS AND
CC       MUSCLE.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN INTERACTS WITH ENDOPHILIN I AND
CC       SORTING NEXIN 9.
CC   -!- DOMAIN: DESINTEGRIN DOMAIN BINDS TO INTEGRIN ALPHAV-BETA3 (BY
CC       SIMILARITY).
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE. AN
CC       ADDITIONAL MEMBRANE PROXIMAL SITE OF CLEAVAGE AFFECTS A SMALL
CC       PERCENTAGE OF THE PROTEIN AND RESULTS IN DISULFIDE-LINKED
CC       FRAGMENTS. THE PRO-DOMAIN IS APPARENTLY CLEAVED IN SEVERAL
CC       POSITIONS THAT ARE N-TERMINAL OF THE FURIN CLEAVAGE SITE.
CC   -!- PTM: MAY BE PARTIALLY SIALYLATED.
CC   -!- PTM: Phosphorylation increases association with PTKs (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to peptidase family M12B.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 disintegrin domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF006196; AAC61896.1; -.
DR   EMBL; AB022089; BAA88903.1; -.
DR   HSSP; P17494; 1KST.
DR   MEROPS; M12.215; -.
DR   MGD; MGI:1333882; Adam15.
DR   InterPro; IPR006586; ADAM_cysteine.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001818; Matrixin.
DR   InterPro; IPR002870; Pep_M12B_propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR006025; Zn_MTpeptdse.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   ProDom; PD000664; Disintegrin; 1.
DR   SMART; SM00608; ACR; 1.
DR   SMART; SM00050; DISIN; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00546; CYSTEINE_SWITCH; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Transmembrane; EGF-like domain; SH3-binding; Phosphorylation.
FT   SIGNAL       1    17       POTENTIAL.
FT   PROPEP      18   207       PROBABLE.
FT   CHAIN      208   815       ADAM 15.
FT   DOMAIN     208   696       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   697   717       POTENTIAL.
FT   DOMAIN     718   815       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     208   415       METALLOPROTEASE.
FT   DOMAIN     422   509       DISINTEGRIN-LIKE.
FT   DOMAIN     510   657       CYS-RICH.
FT   DOMAIN     658   686       EGF-LIKE.
FT   DOMAIN     699   712       POLY-LEU.
FT   SITE       767   773       SH3-BINDING (POTENTIAL).
FT   SITE       802   808       SH3-BINDING (POTENTIAL).
FT   SITE       179   179       CYSTEINE SWITCH.
FT   SITE       289   290       CLEAVAGE (BY FURIN) (POTENTIAL).
FT   METAL      349   349       ZINC (CATALYTIC) (POTENTIAL).
FT   ACT_SITE   350   350       POTENTIAL.
FT   METAL      353   353       ZINC (CATALYTIC) (POTENTIAL).
FT   METAL      359   359       ZINC (CATALYTIC) (POTENTIAL).
FT   DISULFID   324   410       BY SIMILARITY.
FT   DISULFID   481   494       POTENTIAL.
FT   DISULFID   658   668       BY SIMILARITY.
FT   DISULFID   662   674       BY SIMILARITY.
FT   DISULFID   676   685       BY SIMILARITY.
FT   MOD_RES    716   716       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES    736   736       PHOSPHORYLATION (BY SIMILARITY).
FT   CARBOHYD   238   238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   390   390       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   393   393       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   607   607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   612   612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    21    22       PP -> RR (IN REF. 2).
FT   CONFLICT   443   443       E -> Q (IN REF. 2).
FT   CONFLICT   459   459       G -> E (IN REF. 2).
FT   CONFLICT   564   565       SP -> T (IN REF. 2).
FT   CONFLICT   654   654       G -> E (IN REF. 2).
FT   CONFLICT   660   660       R -> S (IN REF. 2).
FT   CONFLICT   703   703       L -> R (IN REF. 2).
FT   CONFLICT   712   712       L -> R (IN REF. 2).
FT   CONFLICT   729   729       L -> R (IN REF. 2).
FT   CONFLICT   797   797       L -> S (IN REF. 2).
FT   CONFLICT   803   805       PAP -> AAS (IN REF. 2).
FT   CONFLICT   810   810       A -> P (IN REF. 2).
```

```
SQ   SEQUENCE   815 AA;  87424 MW;  C064BD3B7347D19B CRC64;

  Query Match             11.2%;  Score 72.5;  DB 1;  Length 815;
  Best Local Similarity   20.9%;  Pred. No. 7.7;
  Matches   28;  Conservative   17;  Mismatches   40;  Indels   49;  Gaps    5;

Qy         14 KKGHGHSY----------------------TTAEEAAGIGILTVILGVLLLIGCWYCRR 50
              :| |||                      |:    |: :  ::| ||:|:|  |  |
Db        660 RKCHGHGVCDSSGHCRCEEGWAPPDCMTQLKATSSLTTGLLLSLLLLLVLVLLGASYWHR 719

Qy         51 RNGYRALMDKSLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAE 110
                :: |       |: |    | ||               :| |  |  |   ::::
Db        720 ARLHQRLCQLK---GSSCQY--RAPQ--------------SCPPERPGPPQRAQQMTGT 759

Qy        111 QS------PPPYSP 118
              :|      |||  |
Db        760 KSQGPTKPPPPRKP 773


RESULT 6
EXPA_DROME
ID   EXPA_DROME     STANDARD;      PRT;  1429 AA.
AC   Q07436;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Expanded protein.
GN   EX.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Imaginal disks;
RX   MEDLINE=94094747; PubMed=8269855;
RA   Boedigheimer M., Laughon A.;
RT   "Expanded: a gene involved in the control of cell proliferation in
RT   imaginal discs.";
RL   Development 118:1291-1301(1993).
RN   [2]
RP   REVISIONS.
RA   Boedigheimer M.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE CONTROL OF CELL PROLIFERATION IN
CC       IMAGINAL DISCS. MAY BIND TO CERTAIN PROTEINS OF SIGNAL
CC       TRANSDUCTION PATHWAYS BY INTERACTION WITH THEIR SH3 DOMAINS.
CC   -!- SUBCELLULAR LOCATION: APICAL SURFACE OF DISC CELLS.
CC   -!- DISEASE: MUTATIONS OF EXPANDED PROTEIN CAUSE HYPERPLASMIA OF THE
CC       IMAGINAL DISC RESULTING IN WING OVERGROWTH. THIS OVERGROWTH IS
CC       LIMITED TO SPECIFIC REGIONS ALONG THE 2 WING AXES. DEFECTS ALSO
CC       IN EYES, HEAD, THORAX AND LIMBS WHERE DUPLICATION AND BULGING
CC       OFTEN OCCUR.
CC   -!- SIMILARITY: Contains 1 FERM domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L14768; AAB39774.1; -.
DR   PIR; T13720; T13720.
DR   FlyBase; FBgn0004583; ex.
DR   InterPro; IPR000299; Band_4.1.
DR   Pfam; PF00373; Band_41; 1.
DR   SMART; SM00295; B41; 1.
DR   PROSITE; PS00660; FERM_1; FALSE_NEG.
DR   PROSITE; PS00661; FERM_2; FALSE_NEG.
DR   PROSITE; PS50057; FERM_3; 1.
KW   Developmental protein; SH3-binding.
FT   DOMAIN       26    399       FERM.
FT   SITE       1008   1016       SH3-BINDING (POTENTIAL).
FT   SITE       1012   1020       SH3-BINDING (POTENTIAL).
FT   SITE       1149   1157       SH3-BINDING (POTENTIAL).
FT   DOMAIN      409    412       POLY-GLU.
FT   DOMAIN      782    788       POLY-PRO.
FT   DOMAIN      952    955       POLY-HIS.
FT   DOMAIN     1002   1005       POLY-PRO.
FT   DOMAIN     1011   1017       POLY-PRO.
FT   DOMAIN     1081   1084       POLY-PRO.
FT   DOMAIN     1149   1154       POLY-PRO.
FT   DOMAIN     1158   1168       POLY-ALA.
FT   DOMAIN     1170   1174       POLY-SER.
FT   DOMAIN     1199   1205       POLY-PRO.
FT   DOMAIN     1416   1424       POLY-GLN.
SQ   SEQUENCE   1429 AA;  153886 MW;  3CB08D2FC4862062 CRC64;

  Query Match             11.2%;  Score 72.5;  DB 1;  Length 1429;
  Best Local Similarity   23.4%;  Pred. No. 14;
  Matches   37;  Conservative   10;  Mismatches   52;  Indels   59;  Gaps    7;

Qy          2 PREDAHFIYGYPKKGHG-------HSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGY 54
              || |::   |   :| |       ||  :|||              |::|      | |
Db        708 PRSDSNVSTGSSFRGDGSDPTDNKHSLLSAEELTN----------LIVG------RGTY 750

Qy         55 --RALMDKSLH-----------VGTQCALTRRCPQEGFDHRDSKVSL---------- 88
                |  :  |||           | :         |  :  |  |  |
Db        751 PSRKTVSSSLHSDCDYVTLPLGDQGEEEVDQPPAPPPPYSARHEKTGLCGPPIAKPIPKP 810

Qy         89 --------QEKNCEPVVPNAP-PAYEKLSAEQSPPPYS 117
                      | | || || ||       : |||||
Db        811 IAVVAPKPDSPPCSPPVPPAPIPAPPPAIRRRDPPPYS 848


RESULT 7
PCH7_HUMAN
ID   PCH7_HUMAN     STANDARD;      PRT;  1069 AA.
AC   O60245; O60246; O60247;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Protocadherin 7 precursor (Brain-heart protocadherin) (BH-Pcdh).
GN   PCDH7 OR BHPCDH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RX   MEDLINE=98277460; PubMed=9615233;
RA   Yoshida K., Yoshitomo-Nakagawa K., Seki N., Sasaki M., Sugano S.;
RT   "Cloning, expression analysis, and chromosomal localization of
RT   BH-protocadherin (PCDH7), a novel member of the cadherin
RT   superfamily.";
RL   Genomics 49:458-461(1998).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=A; Synonyms=BH-Pcdh-a;
CC         IsoId=O60245-1; Sequence=Displayed;
CC       Name=B; Synonyms=BH-Pcdh-b;
CC         IsoId=O60245-2; Sequence=VSP_000704;
CC       Name=C; Synonyms=BH-Pcdh-c;
CC         IsoId=O60245-3; Sequence=VSP_000705, VSP_000706;
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in brain and heart and
CC       at lower levels in various other tissues.
CC   -!- SIMILARITY: Contains 7 cadherin domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB006755; BAA25194.1; -.
DR   EMBL; AB006756; BAA25195.1; -.
DR   EMBL; AB006757; BAA25196.1; -.
DR   PIR; T00041; T00041.
DR   PIR; T00042; T00042.
DR   HSSP; P15116; 1NCJ.
DR   Genew; HGNC:8659; PCDH7.
DR   MIM; 602988; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 6.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 7.
DR   PROSITE; PS00232; CADHERIN_1; 6.
DR   PROSITE; PS50268; CADHERIN_2; 7.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal; Alternative splicing.
FT   SIGNAL       1     28       POTENTIAL.
FT   CHAIN       29   1069       PROTOCADHERIN 7.
FT   DOMAIN      29    879       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   880    900       POTENTIAL.
FT   DOMAIN     901   1069       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      29    143       CADHERIN 1.
FT   DOMAIN     144    308       CADHERIN 2.
FT   DOMAIN     309    415       CADHERIN 3.
FT   DOMAIN     424    535       CADHERIN 4.
FT   DOMAIN     536    639       CADHERIN 5.
FT   DOMAIN     640    742       CADHERIN 6.
FT   DOMAIN     745    862       CADHERIN 7.
FT   CARBOHYD    79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   689    689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   747    747       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   780    780       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   822    822       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   840    840       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   845    845       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC  1059   1069       MRLHPYITVFG -> VRCIPNIFKYPREG (in isoform
FT                               B).
FT                               /FTId=VSP_000704.
FT   VARSPLIC   265    311       Missing (in isoform C).
FT                               /FTId=VSP_000705.
FT   VARSPLIC  1059   1069       MRLHPYITVFG -> PFRRVTFSVVSQPQDPHQGSLQSCYD
FT                               SGLEESETPSSKSSSGPRLGALPLPEDNYERTTPDGSVDSR
FT                               PLPDVALTGKCTRECDEYGHSDSCWMPVRTSPERKKSQPKL
FT                               STFMPVDERGSQEKLANGEAAIMGDRNRNLLNKKLTSSYET
FT                               FSAASFSKNEEANPEDIPLTKTGEYKPSPVNTLTRREVYL
FT                               (in isoform C).
FT                               /FTId=VSP_000706.
SQ   SEQUENCE   1069 AA;  116104 MW;  315349BE3414F66D CRC64;

  Query Match             10.9%;  Score 71;  DB 1;  Length 1069;
  Best Local Similarity   25.5%;  Pred. No. 15;
  Matches   27;  Conservative   14;  Mismatches   31;  Indels   34;  Gaps    5;

Qy         20 SYTTAEEAAGI------GILTVILGVLLLIGCWYCR--RRNGYRALMDKSLHVGTQCALT 71
              ||  :::   |      ||:|||| :|:::   |||   :||| |       | :
Db        870 SYEISKQRLSIVIGVVAGIMTVILIILIVVMARYCRSKNKNGYEA--------GKK---- 917

Qy         72 RRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYS 117
                 || |     |         :  |  :| : :   | ||
Db        918 --------DHEDFFTPQQHDK------SKKPKKDKKNKKSKQPLYS 949


RESULT 8
TMG2_HUMAN
ID   TMG2_HUMAN     STANDARD;      PRT;   202 AA.
AC   O14669;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transmembrane gamma-carboxyglutamic acid protein 2 precursor (Proline-
DE   rich Gla protein 2) (Proline-rich gamma-carboxyglutamic acid protein
DE   2).
GN   PRRG2 OR TMG2 OR PRGP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97404347; PubMed=9256434;
RA   Kulman J.D., Harris J.E., Haldeman B.A., Davie E.W.;
RT   "Primary structure and tissue distribution of two novel proline-rich
RT   gamma-carboxyglutamic acid proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:9058-9062(1997).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: Highly expressedd in the thyroid.
CC   -!- PTM: Gla residues are produced after subsequent posttranslational
CC       modifications of glutamic acid by a vitamin K-dependent gamma-
CC       carboxylase.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF009243; AAB67071.1; -.
DR   HSSP; P00740; 1CFH.
DR   Genew; HGNC:9470; PRRG2.
DR   MIM; 604429; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00594; gla; 1.
DR   SMART; SM00069; GLA; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
KW   Gamma-carboxyglutamic acid; Vitamin K; Transmembrane; Signal.
FT   SIGNAL       1     23       POTENTIAL.
FT   PROPEP      24     49
FT   CHAIN       50    202       TRANSMEMBRANE GAMMA-CARBOXYGLUTAMIC ACID
FT                               PROTEIN 2.
FT   DOMAIN      50    109       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   110    132       POTENTIAL.
FT   DOMAIN     133    202       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      54     91       GLA-RICH.
FT   DOMAIN     161    173       POLY-PRO.
FT   DOMAIN     191    194       POLY-PRO.
SQ   SEQUENCE   202 AA;  22393 MW;  BC79400C98492060 CRC64;

  Query Match             10.9%;  Score 70.5;  DB 1;  Length 202;
  Best Local Similarity   29.7%;  Pred. No. 2.8;
  Matches   38;  Conservative    8;  Mismatches   37;  Indels   45;  Gaps   10;

Qy         10 YGYPKKGHGHSYTTAEEAAGI--GILTVILGVLLLIGC-WYCRRRNGYRALMDKSLHVGT 66
              | |  ||       |  | |:  ||| ::|  |   |  || | |           | |
Db         94 YIYNGKGGRGRVDVASLAVGLTGGILLIVLAGL---GAFWYLRWRQ----------HRGQ 140

Qy         67 QCALTRRCPQEGFDHRDSKVS-LQEKNCEPVVPNAP----------PAYEKLSA----- 109
              |     ||||       :| |   |  |: | |               | ||:  |
Db        141 Q-----PCPQEA-----GLISPLSPLN--PLGPPTPLPPPPPPPPPGLPTYEQALAASGVH 188

Qy        110 EQSPPPYS 117
              :  ||||:
Db        189 DAPPPPYT 196


RESULT 9
TMG1_HUMAN
ID   TMG1_HUMAN     STANDARD;      PRT;   218 AA.
AC   O14668;
```

```
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transmembrane gamma-carboxyglutamic acid protein 1 precursor (Proline-
DE   rich Gla protein 1) (Proline-rich gamma-carboxyglutamic acid protein
DE   1).
GN   PRRG1 OR TMG1 OR PRGP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97404347; PubMed=9256434;
RA   Kulman J.D., Harris J.E., Haldeman B.A., Davie E.W.;
RT   "Primary structure and tissue distribution of two novel proline-rich
RT   gamma-carboxyglutamic acid proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:9058-9062(1997).
CC   -!- TISSUE SPECIFICITY: Highly expressed in the spinal cord.
CC   -!- PTM: Gla residues are produced after subsequent posttranslational
CC       modifications of glutamic acid by a vitamin K-dependent gamma-
CC       carboxylase.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF009242; AAB67070.1; -.
DR   HSSP; P00740; 1CFH.
DR   Genew; HGNC:9469; PRRG1.
DR   MIM; 604428; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00594; gla; 1.
DR   PRINTS; PR00001; GLABLOOD.
DR   SMART; SM00069; GLA; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
KW   Gamma-carboxyglutamic acid; Vitamin K; Transmembrane.
FT   PROPEP          1     20       POTENTIAL.
FT   CHAIN          21    218       TRANSMEMBRANE GAMMA-CARBOXYGLUTAMIC ACID
FT                                  PROTEIN 1.
FT   DOMAIN         21     83       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM       84    106       POTENTIAL.
FT   DOMAIN        107    218       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN         24     61       GLA-RICH.
FT   DOMAIN        131    135       POLY-PRO.
SQ   SEQUENCE   218 AA;  24947 MW;  26538A61AB0AEB98 CRC64;

  Query Match             10.7%;  Score 69.5;  DB 1;  Length 218;
  Best Local Similarity   27.0%;  Pred. No. 3.9;
  Matches   30;  Conservative    9;  Mismatches   37;  Indels   35;  Gaps    5;

Qy         33 LTVILGVLLLIGCWYCRRRNGYRALMDKSLHVGTQCALT----RRCPQEGFDH------- 81
              | :|| |: |:  | |  ||  |       |:     |        | | ||
Db         91 LFIILLVIFLI--WRCFLRNKTRRQTVTEGHIPFPQHLNIITPPPPPDEVFDSSGLSPGF 148

Qy         82 ------RDSKVSLQEKNCEPVVPNAPPAYEKLSAEQS----------PPP 115
                  |   || :  ||:|     || ||: : : :           |||
Db        149 LGYVVGRSDSVSTRLSNCDP-----PPTYEEATGQVNLQRSETEPHLDPPP 194


RESULT 10
LSG1_HAEIN
ID   LSG1_HAEIN     STANDARD;      PRT;   401 AA.
AC   P71399; Q48210;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Lsg locus putative protein 1.
GN   HI1700.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A2;
RA   McLaughlin R., Abu Kwaik Y., Young R., Spinola S., Apicella M.;
RT   "Characterization and sequence of the lsg locus from Haemophilus
RT   influenzae.";
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the polysaccharide synthase family.
CC       HI0867/HI1700 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M94855; AAA24978.1; -.
DR   EMBL; U32842; AAC23346.1; ALT_INIT.
DR   PIR; H64175; H64175.
DR   TIGR; HI1700; -.
DR   InterPro; IPR002797; Polysacc_synt.
DR   Pfam; PF01943; Polysacc_synt; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM       8     28       POTENTIAL.
FT   TRANSMEM      36     56       POTENTIAL.
FT   TRANSMEM      87    107       POTENTIAL.
FT   TRANSMEM     132    152       POTENTIAL.
FT   TRANSMEM     162    182       POTENTIAL.
FT   TRANSMEM     199    219       POTENTIAL.
FT   TRANSMEM     237    257       POTENTIAL.
FT   TRANSMEM     282    302       POTENTIAL.
FT   TRANSMEM     320    340       POTENTIAL.
FT   TRANSMEM     352    372       POTENTIAL.
FT   TRANSMEM     374    394       POTENTIAL.
FT   CONFLICT     249    249       V -> I (IN REF. 1).
FT   CONFLICT     276    276       V -> I (IN REF. 1).
FT   CONFLICT     358    358       I -> V (IN REF. 1).
SQ   SEQUENCE   401 AA;  45944 MW;  FE2E7B02747B0874 CRC64;

  Query Match             10.7%;  Score 69.5;  DB 1;  Length 401;
  Best Local Similarity   21.6%;  Pred. No. 7.5;
  Matches   21;  Conservative   20;  Mismatches   23;  Indels   33;  Gaps    4;

Qy          6 AHFIYGYPKK------GHGHSYTTAEEAAGIGILTVILGVLLLIGCW--------YCRRR 51
              : : | | |:        |::|||            |:| ::|||||        |
Db         63 SRYFYFYGKRSLNLVVNTGYAYTT------------IIGSIILIGCWIAQSEILFYAALS 110

Qy         52 NGYRALMDKSLHVGTQCALTRRCPQEGFDHRDSKVSL 88
```

```
              : ::: ::  | |       |:| :: : :   : ||
Db        111 SIFQSFLNVQLSV-------RQCQKKAWSYAFIQFSL 140


RESULT 11
CYY1_MOUSE
ID   CYY1_MOUSE     STANDARD;      PRT;   165 AA.
AC   Q8VIH7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cysteine and tyrosine-rich protein 1 precursor (Proline-rich domain
DE   containing protein).
GN   CYYR1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=22057895; PubMed=12062809;
RA   Vitale L., Casadei R., Canaider S., Lenzi L., Strippoli P.,
RA   D'Addabbo P., Giannone S., Carinci P., Zannotti M.;
RT   "Cysteine and tyrosine-rich 1 (CYYR1), a novel unpredicted gene on
RT   human chromosome 21 (21q21.2), encodes a cysteine and tyrosine-rich
RT   protein and defines a new family of highly conserved vertebrate-
RT   specific genes.";
RL   Gene 290:141-151(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22032984; PubMed=12036297;
RA   Reymond A., Camargo A.A., Deutsch S., Stevenson B.J., Parmigiani R.B.,
RA   Ucla C., Bettoni F., Rossier C., Lyle R., Guipponi M., de Souza S.,
RA   Iseli C., Jongeneel C.V., Bucher P., Simpson A.J.G.,
RA   Antonarakis S.E.;
RT   "Nineteen additional unpredicted transcripts from human chromosome
RT   21.";
RL   Genomics 79:824-832(2002).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF442733; AAL35295.1; -.
DR   EMBL; AY061854; AAL35738.1; -.
DR   MGD; MGI:2152187; Cyyr1.
KW   Signal; Transmembrane.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30    165       CYSTEINE AND TYROSINE-RICH PROTEIN 1.
FT   DOMAIN       30     61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     62     82       POTENTIAL.
FT   DOMAIN       83    165       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      144    149       POLY-PRO.
FT   CONFLICT     36     36       R -> H (IN REF. 2).
SQ   SEQUENCE   165 AA;  18063 MW;  9494D0AAE2BCFDC1 CRC64;

  Query Match             10.5%;  Score 68;  DB 1;  Length 165;
  Best Local Similarity   21.2%;  Pred. No. 4.2;
  Matches   22;  Conservative   19;  Mismatches   37;  Indels   26;  Gaps    5;

Qy         28 AGIGILTVILGVLLLIG---------CWYCRRRNGYRALMDKSLHVGTQCALTRRCPQEG 78
              :|  |  :: |:: ::|         |   :   | |  : :: |:    :     |
Db         59 SGTAIAGIVFGIVFIMGVIAGIAICICMCMKNNRGTRVGVIRAAHINA-ISYPMAPPPYT 117

Qy         79 FDHR-DSKVSLQEKNCEPVVPNAPPAYE---KLSAEQSPPPYSP 118
              :||  : :  |           || |    : ||::|||  |
Db        118 YDHEMEYRTDL-----------PPPYSAAPQASAQRSPPPYP 149


RESULT 12
COME_METJA
ID   COME_METJA     STANDARD;      PRT;   188 AA.
AC   P58416; Q57704;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Sulfopyruvate decarboxylase subunit beta (EC 4.1.1.79).
GN   COME OR MJ0256.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=20398170; PubMed=10940029;
RA   Graupner M., Xu H., White R.H.;
RT   "Identification of the gene encoding sulfopyruvate decarboxylase, an
RT   enzyme involved in biosynthesis of coenzyme M.";
RL   J. Bacteriol. 182:4862-4867(2000).
CC   -!- FUNCTION: Catalyzes the decarboxylation of sulfopyruvic acid to
CC       sulfoacetaldehyde.
CC   -!- CATALYTIC ACTIVITY: 3-sulfopyruvate = 2-sulfoacetaldehyde + CO(2).
CC   -!- COFACTOR: TPP (Potential).
CC   -!- PATHWAY: Coenzyme M biosynthesis; fourth step.
CC   -!- SUBUNIT: Heterododecamer composed of 6 subunits alpha and 6
CC       subunits beta.
CC   -!- MISCELLANEOUS: Inactivated by oxygen when heated in air at 80
CC       degrees Celsius.
CC   -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC   -!- CAUTION: The sequence corresponding to this entry was originally
CC       entered in SWISS-PROT as AC Q57704 in November 1997 and was
CC       deleted in July 1999 because TIGR removed the CDS for that ORF. We
CC       have recreated it because of the evidence (Ref.2) that it really
CC       exists.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67480; -; NOT_ANNOTATED_CDS.
DR   PIR; A64332; A64332.
DR   TIGR; MJ0256; -.
DR   InterPro; IPR000399; Pyruvate_decarb.
DR   Pfam; PF02775; TPP_enzymes_C; 1.
DR   PROSITE; PS00187; TPP_ENZYMES; FALSE_NEG.
KW   Coenzyme M biosynthesis; Lyase; Decarboxylase; Thiamine pyrophosphate;
KW   Complete proteome.
SQ   SEQUENCE   188 AA;  20980 MW;  52A6C91E2D43B97D CRC64;

  Query Match             10.4%;  Score 67.5;  DB 1;  Length 188;
  Best Local Similarity   24.8%;  Pred. No. 5.4;
```

```
  Matches   26;  Conservative   13;  Mismatches   45;  Indels   21;  Gaps    3;

Qy         29 GIGILTVILGVLLLIGCWYCRRRNGYRALMDKS---------LHVGTQCALTRRCPQEGF 79
              | | : : || |  ||  |   :|    ::| |         | |    |       |
Db         72 GDGSILMNLGSLSTIG--YMNPKNYILVIIDNSAYGSTGNQKTHTGKNTNLEEIAKGCGL 129

Qy         80 D----------HRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPP 114
              |           :: | :| |: |: ::    |  || |  : ||
Db        130 DTITTESLEEFEKEFKNALNEEKCKVIIAKTIPYNEKCSNIEIPP 174


RESULT 13
MGD1_HUMAN
ID   MGD1_HUMAN     STANDARD;      PRT;   778 AA.
AC   Q9Y5V3; Q9H352; Q9HBT4; Q9UF36;
DT   16-OCT-2001 (Rel. 40, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Melanoma-associated antigen D1 (MAGE-D1 antigen) (Neurotrophin
DE   receptor-interacting MAGE homolog) (PRO2292).
GN   MAGED1 OR NRAGE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=99339980; PubMed=10409427;
RA   Pold M., Zhou J., Chen G.L., Hall J.M., Vescio R.A., Berenson J.R.;
RT   "Identification of a new, unorthodox member of the MAGE gene family.";
RL   Genomics 59:161-167(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20439481; PubMed=10985348;
RA   Salehi A.H., Roux P.P., Kubu C.J., Zeindler C., Bhakar A.,
RA   Tannis L.-L., Verdi J.M., Barker P.A.;
RT   "NRAGE, a novel MAGE protein, interacts with the p75 neurotrophin
RT   receptor and facilitates nerve growth factor dependent apoptosis.";
RL   Neuron 27:279-288(2000).
RN   [3]
RP   SEQUENCE OF 304-778 FROM N.A.
RC   TISSUE=Testis;
RA   Blum H., Bauersachs S., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 396-778 FROM N.A.
RC   TISSUE=Fetal liver;
RX   MEDLINE=21177478; PubMed=11280991;
RA   Zhang C.G., Xing G.C., Wei H.D., Yu Y.T., He F.C.;
RT   "A new melanoma antigen-encoding gene subfamily in human chromosome
RT   X.";
RL   I Chuan Hsueh Pao 28:197-203(2001).
RN   [5]
RP   IDENTIFICATION OF THE TRANSLATIONAL INITIATION CODON.
RX   MEDLINE=20541720; PubMed=11087672;
RA   Kubu C.J., Goldhawk D.G., Barker P.A., Verdi J.M.;
RT   "Identification of the translational initiation codon in human
RT   MAGED1.";
RL   Genomics 70:150-152(2000).
CC   -!- FUNCTION: Involved in the apoptotic response after nerve growth
CC       factor (NGF) binding in neuronal cells. Binds p75NTR and
CC       antagonizes its association with TrkA, inhibits cell cycle
CC       progression, and facilitates p75NTR-mediated apoptosis. May act as
CC       a regulator of the function of DLX family members.
CC   -!- SUBUNIT: Interacts with the p75 neurotrophine receptor.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. Expression shifts from the
CC       cytoplasm to the plasma membrane upon stimulation with NGF (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BONE MARROW STROMAL CELLS FROM
CC       BOTH MULTIPLE MYELOMA PATIENTS AND HEALTHY DONORS. SEEMS TO BE
CC       UBIQUITOUSLY EXPRESSED.
CC   -!- SIMILARITY: Contains 1 MAGE domain.
CC   -!- CAUTION: REF.1 differs from that shown due to several frameshifts
CC       that resulted in a N-terminally truncated protein.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF124440; AAD31421.1; ALT_FRAME.
DR   EMBL; AF217963; AAG09704.1; -.
DR   EMBL; AL133628; CAB63752.1; -.
DR   EMBL; AF132205; AAG35551.1; ALT_INIT.
DR   Genew; HGNC:6813; MAGED1.
DR   MIM; 300224; -.
DR   GO; GO:0008222; F:tumor antigen; TAS.
DR   InterPro; IPR002190; MAGE.
DR   Pfam; PF01454; MAGE; 1.
DR   PROSITE; PS50838; MAGE; 1.
KW   Antigen; Multigene family; Repeat.
FT   DOMAIN      296    444       22 X 6 AA TANDEM REPEATS OF W-[PQ]-X-P-X-
FT                                X.
FT   REPEAT      296    301       1.
FT   REPEAT      302    307       2.
FT   REPEAT      308    313       3.
FT   REPEAT      332    337       4.
FT   REPEAT      338    343       5.
FT   REPEAT      344    349       6.
FT   REPEAT      350    355       7.
FT   REPEAT      356    361       8.
FT   REPEAT      362    367       9.
FT   REPEAT      368    373       10.
FT   REPEAT      374    379       11.
FT   REPEAT      380    385       12.
FT   REPEAT      386    391       13.
FT   REPEAT      392    397       14.
FT   REPEAT      398    403       15.
FT   REPEAT      404    409       16.
FT   REPEAT      410    415       17.
FT   REPEAT      416    421       18.
FT   REPEAT      422    427       19.
FT   REPEAT      428    432       20 (IMPERFECT).
FT   REPEAT      433    438       21.
FT   REPEAT      439    444       22.
FT   DOMAIN      471    669       MAGE.
SQ   SEQUENCE   778 AA;  86150 MW;  0F8BEC7155326FCC CRC64;

  Query Match             10.4%;  Score 67.5;  DB 1;  Length 778;
  Best Local Similarity   29.7%;  Pred. No. 25;
  Matches   33;  Conservative   10;  Mismatches   33;  Indels   35;  Gaps    6;

Qy         19 HSY---TTAEEAAGI----------GILTVILGVLLLIG-------CWYCRRRNGYRALM 58
              | |    :| |  |||          |:| |||||: : |       |   |: | |  :
Db        536 HLYILISTPESLAGILGTTKDTPKLGLLLVILGVIFMNGNRASEAVLWEALRKMGLRPGV 595

Qy         59 DKSLHVGTQCALTRR-CPQEGFDHRDSKVSLQEKNCEPVVPNA-PPAYEKL 107
                 |    :  ||     |: |:|             |||: || || |
Db        596 RHPLLGDLRKLLTYEFVKQKYLDYRR-----------VPNSNPPEYEFL 633


RESULT 14
FATH_HUMAN
ID   FATH_HUMAN     STANDARD;      PRT;  4590 AA.
AC   Q14517;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cadherin-related tumor suppressor homolog precursor (Fat protein
DE   homolog).
GN   FAT.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymphocytes;
RX   MEDLINE=96163873; PubMed=8586420;
RA   Dunne J., Hanby A.M., Poulsom R., Jones T.A., Sheer D., Chin W.G.,
RA   Da S.M., Zhao Q., Beverley P.C.L., Owen M.J.;
RT   "Molecular cloning and tissue expression of FAT, the human homologue
RT   of the Drosophila fat gene that is located on chromosome 4q34-q35 and
RT   encodes a putative adhesion molecule.";
RL   Genomics 30:207-223(1995).
CC   -!- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MANY EPITHELIAL AND SOME
CC       ENDOTHELIAL AND SMOOTH MUSCLE CELLS.
CC   -!- SIMILARITY: Contains 34 cadherin domains.
CC   -!- SIMILARITY: Contains 5 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 laminin G-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X87241; CAA60685.1; -.
DR   HSSP; P01132; 1EGF.
DR   Genew; HGNC:3595; FAT.
DR   MIM; 600976; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0008181; F:tumor suppressor; TAS.
DR   GO; GO:0007155; P:cell adhesion; TAS.
DR   GO; GO:0007267; P:cell-cell signaling; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00028; cadherin; 33.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 31.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00232; CADHERIN_1; 16.
DR   PROSITE; PS50268; CADHERIN_2; 33.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Cell adhesion; Signal; Glycoprotein; Transmembrane; Calcium-binding;
KW   Repeat; EGF-like domain.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   4590       CADHERIN-RELATED TUMOR SUPPRESSOR
FT                               HOMOLOG.
FT   DOMAIN      22   4183       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  4184   4204       POTENTIAL.
FT   DOMAIN    4205   4590       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      22    149       CADHERIN 1.
FT   DOMAIN     150    256       CADHERIN 2.
FT   DOMAIN     257    361       CADHERIN 3.
FT   DOMAIN     362    463       CADHERIN 4.
FT   DOMAIN     464    569       CADHERIN 5.
FT   DOMAIN     570    716       CADHERIN 6.
FT   DOMAIN     717    822       CADHERIN 7.
FT   DOMAIN     823    927       CADHERIN 8.
FT   DOMAIN     928   1034       CADHERIN 9.
FT   DOMAIN    1035   1138       CADHERIN 10.
FT   DOMAIN    1139   1245       CADHERIN 11.
FT   DOMAIN    1246   1345       CADHERIN 12.
FT   DOMAIN    1346   1456       CADHERIN 13.
FT   DOMAIN    1457   1562       CADHERIN 14.
FT   DOMAIN    1563   1670       CADHERIN 15.
FT   DOMAIN    1671   1769       CADHERIN 16.
FT   DOMAIN    1770   1882       CADHERIN 17.
FT   DOMAIN    1883   1982       CADHERIN 18.
FT   DOMAIN    1983   2084       CADHERIN 19.
FT   DOMAIN    2085   2185       CADHERIN 20.
FT   DOMAIN    2186   2286       CADHERIN 21.
FT   DOMAIN    2287   2393       CADHERIN 22.
FT   DOMAIN    2394   2495       CADHERIN 23.
FT   DOMAIN    2496   2599       CADHERIN 24.
FT   DOMAIN    2600   2705       CADHERIN 25.
FT   DOMAIN    2706   2811       CADHERIN 26.
FT   DOMAIN    2812   2920       CADHERIN 27.
FT   DOMAIN    2921   3024       CADHERIN 28.
FT   DOMAIN    3025   3127       CADHERIN 29.
FT   DOMAIN    3128   3232       CADHERIN 30.
FT   DOMAIN    3233   3337       CADHERIN 31.
FT   DOMAIN    3338   3442       CADHERIN 32.
FT   DOMAIN    3443   3546       CADHERIN 33.
FT   DOMAIN    3547   3649       CADHERIN 34.
FT   DOMAIN    3790   3827       EGF-LIKE 1.
FT   DOMAIN    3831   4011       LAMININ G-LIKE.
FT   DOMAIN    4013   4051       EGF-LIKE 2.
FT   DOMAIN    4052   4089       EGF-LIKE 3.
FT   DOMAIN    4090   4126       EGF-LIKE 4.
FT   DOMAIN    4127   4163       EGF-LIKE 5.
FT   CARBOHYD    40     40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   660    660       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   740    740       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   791    791       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   998    998       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1426   1426       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1551   1551       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1751   1751       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1867   1867       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1905   1905       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1943   1943       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1994   1994       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  2328   2328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  2467   2467       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3326   3326       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3424   3424       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3446   3446       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3615   3615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3642   3642       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3718   3718       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4154   4154       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   4590 AA;  506273 MW;  04483CCD000E60A7 CRC64;

  Query Match             10.4%;  Score 67.5;  DB 1;  Length 4590;
  Best Local Similarity   27.7%;  Pred. No. 1.6e+02;
  Matches   31;  Conservative   14;  Mismatches   38;  Indels   29;  Gaps    6;

Qy         27 AAGIGILTVILGVLLLIGCW-YCRR---RNGYRALMDKSLHVGTQCALTRRCPQEGFDHR 82
              | ||||:  : |: ||:  :  ||:   |       |  |:|   |  :|     :
Db       4182 AEGIGIVVFVAGIFLLVVVFVLCRKMISRKKKHQAEPKDKHLGPATAFLQR------PYF 4235

Qy         83 DSKVSLQEKN----CEPVVPNAPPAY-----------EKLSAEQSPPPYSP 118
              |||::   ||      | ||  | :|           :: | | |  |  |
Db       4236 DSKLN---KNIYSDIPPQVPVRPISYTPSIPSDSRNNLDRNSFEGSAIPEHP 4284


RESULT 15
MGD1_MOUSE
ID   MGD1_MOUSE     STANDARD;      PRT;   775 AA.
```

```
AC   Q9QYH6; Q99PB5; Q9CYX1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Melanoma-associated antigen D1 (MAGE-D1 antigen) (Neurotrophin
DE   receptor-interacting MAGE homolog) (Dlxin-1).
GN   MAGED1 OR NRAGE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=21265065; PubMed=11084035;
RA   Masuda Y., Sasaki A., Shibuya H., Ueno N., Ikeda K., Watanabe K.;
RT   "Dlxin-1, a novel protein that binds Dlx5 and regulates its
RT   transcriptional function.";
RL   J. Biol. Chem. 276:5331-5338(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Aujuier P.H., Chomez P.M., De Backer O.R., Bertrand M.J.M.;
RT   "Ten new murine members of the MAGE gene family.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: Involved in the apoptotic response after nerve growth
CC       factor (NGF) binding in neuronal cells. Binds p75NTR and
CC       antagonizes its association with TrkA, inhibits cell cycle
CC       progression, and facilitates p75NTR-mediated apoptosis. May act as
CC       a regulator of the function of DLX family members (By similarity).
CC   -!- SUBUNIT: INTERACTS WITH DLX5, DLX7 AND MSX2 AND FORMS
CC       HOMOMULTIMERS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. Expression shifts from the
CC       cytoplasm to the plasma membrane upon stimulation with NGF (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed in many adult tissues,
CC       except for the spleen. Expressed in osteoblastic and
CC       chondrogenic cell lines and also during embryonic development.
CC   -!- SIMILARITY: Contains 1 MAGE domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB029448; BAA87959.1; -.
DR   EMBL; AF319975; AAK01203.1; -.
DR   EMBL; AK017275; BAB30666.1; -.
DR   EMBL; AK013231; -; NOT_ANNOTATED_CDS.
DR   MGD; MGI:1930187; Maged1.
DR   GO; GO:0005515; F:protein binding activity; IPI.
DR   GO; GO:0003713; F:transcription co-activator activity; IDA.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; IDA.
DR   InterPro; IPR002190; MAGE.
DR   Pfam; PF01454; MAGE; 1.
DR   PROSITE; PS50838; MAGE; 1.
KW   Antigen; Multigene family; Repeat.
FT   DOMAIN      292    441       22 X 6 AA TANDEM REPEATS OF W-[PQ]-X-P-X-
FT                                X.
FT   REPEAT      292    297       1.
FT   REPEAT      298    303       2.
FT   REPEAT      304    309       3.
FT   REPEAT      329    334       4.
FT   REPEAT      335    340       5.
FT   REPEAT      341    346       6.
FT   REPEAT      347    352       7.
FT   REPEAT      353    358       8.
FT   REPEAT      359    364       9.
FT   REPEAT      365    370       10.
FT   REPEAT      371    376       11.
FT   REPEAT      377    382       12.
FT   REPEAT      383    388       13.
FT   REPEAT      389    394       14.
FT   REPEAT      395    400       15.
FT   REPEAT      401    406       16.
FT   REPEAT      407    412       17.
FT   REPEAT      413    418       18.
FT   REPEAT      419    424       19.
FT   REPEAT      425    429       20 (IMPERFECT).
FT   REPEAT      430    435       21.
FT   REPEAT      436    441       22.
FT   DOMAIN      468    666       MAGE.
FT   CONFLICT    357    362       MISSING (IN REF. 2).
SQ   SEQUENCE   775 AA;  85669 MW;  224B82470816835A CRC64;

  Query Match             10.2%;  Score 66.5;  DB 1;  Length 775;
  Best Local Similarity   28.8%;  Pred. No. 31;
  Matches   32;  Conservative   11;  Mismatches   33;  Indels   35;  Gaps    6;

Qy         19 HSY---TTAEEAAGI----------GILTVILGVLLLIG-------CWYCRRRNGYRALM 58
              | |   :| |  |||          |:| ||||:: : |       |   |: | |  :
Db        533 HLYILISTPESLAGILGTTKDTPKLGLLLVILGIIFMNGNRATEAVLWEALRKMGLRPGV 592

Qy         59 DKSLHVGTQCALTRR-CPQEGFDHRDSKVSLQEKNCEPVVPNA-PPAYEKL 107
                 |    :  ||     |:  |:|              |||: || || |
Db        593 RHPLLGDLRKLLTYEFVKQKYLDYRR------------VPNSNPPEYEFL 630


Search completed: October  7, 2003, 18:48:34
Job time : 13 secs
```